

# Fixed-Smoothing Asymptotics For Time Series

Xianyang Zhang and Xiaofeng Shao\*

University of Illinois at Urbana-Champaign

March 2, 2013

**Abstract** In this paper, we propose a class of estimators for estimating the asymptotic covariance matrix of the generalized method of moments estimator in the stationary time series models. Our proposal provides a unification of the existing smoothing parameter dependent covariance estimators, including the traditional heteroskedasticity and autocorrelation consistent covariance estimator and some recently developed estimators, such as cluster-based covariance estimator and projection-based covariance estimator. Under mild conditions, we establish the first order asymptotic distribution for the Wald statistics when the smoothing parameter is held fixed. Furthermore, we derive higher order Edgeworth expansions for the finite sample distribution of the Wald statistics in the Gaussian location model under the fixed-smoothing paradigm. In particular, we show that the error of asymptotic approximation is at the order of the reciprocal of the sample size and obtain explicit forms for the leading error terms in the expansions. The results are used to justify the second order correctness of a new bootstrap method, the Gaussian dependent bootstrap, in the context of Gaussian location model. Some simulation results are also presented to corroborate our theoretical findings.

**Keywords:** Bootstrap; Fixed-smoothing asymptotics; Generalized method of moments; Higher order expansion; Long run variance matrix.

**JEL Classification Numbers:** C12, C22

---

\*Address: Department of Statistics, University of Illinois at Urbana-Champaign, 725 South Wright St, Champaign, IL, 61820, United States. Emails: zhang104@illinois.edu, xshao@illinois.edu. Shao acknowledges partial research support from the NSF under Grant No. DMS08-04937 and DMS11-04545.

# 1 Introduction

Many economic and financial applications involve time series data with autocorrelation and heteroskedasticity properties. Often the unknown dependence structure is not the chief object of interest but the inference on the parameter of interest involves the estimation of unknown dependence. In stationary time series models estimated by generalized method of moments (GMM), robust inference is typically accomplished by consistently estimating the asymptotic covariance matrix, which is proportional to the long run variance (LRV) matrix of the estimating equations or moment conditions defining the estimator, using a kernel smoothing method. In the econometric and statistical literature, the bandwidth parameter/truncation lag involved in the kernel smoothing method is assumed to grow slowly with sample size in order to achieve consistency. The inference is conducted by plugging in a covariance matrix estimator that is consistent under heteroskedasticity and autocorrelation. This approach dates back to Newey and West (1987) and Andrews (1991). Recently, Kiefer and Vogelsang (2005) (KV, hereafter) developed an alternative first order asymptotic theory for the HAC (heteroskedasticity and autocorrelation consistent) based robust inference, where the proportion of the bandwidth involved in the HAC estimator to the sample size  $T$ , denoted as  $b$ , is held fixed in the asymptotics. Under the fixed- $b$  asymptotics, the HAC estimator converges to a nondegenerate yet nonstandard limiting distribution. The tests based on the fixed- $b$  asymptotic approximation were shown to enjoy better finite sample properties than the tests based on the small- $b$  asymptotic theory under which the HAC estimator is consistent and the limiting distribution of the studentized statistic admits a standard form, such as standard normal or  $\chi^2$  distribution. Using the higher-order Edgeworth expansions, Jansson (2004), Sun et al. (2008) and Sun (2010) rigorously proved that the fixed- $b$  asymptotics provides a high order refinement over the traditional small- $b$  asymptotics in the Gaussian location model. Sun et al. (2008) also provided an interesting decision theoretical justification for the use of fixed- $b$  rules in econometric testing. For non-Gaussian linear processes, Gonçalves and Vogelsang (2011) obtained an upper bound on the convergence rate of the error in the fixed- $b$  approximation and showed that it can be smaller than the error of the normal approximation under suitable assumptions.

Since the seminal contribution by KV (2005), there has been a growing body of work in econometrics and statistics to extend and expand the fixed- $b$  idea in the inference for time series data. For example, Sun (2011a) developed a procedure for hypothesis testing in time series models by using the nonparametric series method. The basic idea is to project the time series onto a space spanned by a set of fourier basis functions [see Phillips (2005) for an early development] and construct the covariance matrix estimator based on the projection vectors with the number of basis functions held fixed. Also see Sun (2011b) for the use of a similar idea in the inference of the trend regression models. Ibragimov and Müller (2010) proposed a subsampling based  $t$ -statistic for robust inference where the unknown dependence structure can be in the temporal, spatial or other forms. In their paper, the number of non-overlapping blocks is held fixed. The  $t$ -statistic based approach was extended by Bester et al. (2009) to the inference of spatial and panel data with group structure. In the context of misspecification testing, Chen

and Qu (2010) proposed a modified  $M$  test of Kuan and Lee (2006) which involves dividing the full sample into several recursive subsamples and constructing a normalization matrix based on them. In the statistical literature, Shao (2010) developed the self-normalized approach to inference for time series data that uses an inconsistent long run variance estimator based on recursive subsample estimates. The self-normalized method is an extension of Lobato (2001) from the sample autocovariances to more general approximately linear statistics and it coincides with KV's fixed- $b$  approach in the inference of the mean of a stationary time series by using the Bartlett kernel and letting  $b = 1$ . Although the above inference procedures are proposed in different settings and for different problems and data structure, they share a common feature in the sense that the underlying smoothing parameters in the asymptotic covariance matrix estimator such as the number of basis functions, the number of cluster groups and the number of recursive subsamples, play a similar role as the bandwidth in the HAC estimator. Throughout the paper, we shall call these asymptotics, where the smoothing parameter (or function of smoothing parameter) is held fixed, the fixed-smoothing asymptotics. In contrast, when the smoothing parameter grows with respect to sample size, we use the term increasing-domain asymptotics. At some places the terms fixed- $K$  (or fixed- $b$ ) and increasing- $K$  (or small- $b$ ) asymptotics are used to follow the convention in the literature.

In this article, we make several methodological and theoretical contributions to the fixed-smoothing literature. First, we propose a general class of estimators for estimating the LRV matrix in the inference of stationary time series models estimated by GMM. Our proposal includes the traditional lag window type (or HAC) covariance estimator, the projection-based covariance estimator, the cluster-based covariance estimator and the blockwise recursive subsampling-based covariance estimator as special cases. The general covariance estimator considered here involves projecting the original data onto a space spanned by a sequence of basis functions (not necessarily orthogonal), where the number of basis functions  $K$  plays a key role in determining asymptotic properties of the estimator. Under the fixed- $K$  asymptotics, we show that the Wald statistic based on the general LRV estimator converges to an (approximate)  $F$  distribution with a scale constant depending only on  $K$  and the number of restrictions being tested. Thus our result provides a unification of the various recently proposed fixed-smoothing inference procedures in the first order sense.

Second, we derive a higher order expansion of the distribution of subsampling  $t$ -statistic when the underlying smoothing parameter  $K$  is held fixed, under the framework of the Gaussian location model. Specifically, we show that the error in the rejection probability (ERP, hereafter) is of order  $O(1/T)$  under the fixed- $K$  asymptotics. These results are similar to those obtained under the fixed- $b$  asymptotics [see Sun et al. (2008)], but are stronger in the sense that we are able to derive the exact form of the leading error term with order  $O(1/T)$ . Building on the new technical arguments used in proving expansion results for the subsampling  $t$ -statistic, we further study the expansion of the distribution of the Wald statistic with the HAC covariance estimator. Under the assumption that the eigenfunctions of the kernel in the HAC estimator have zero mean and other mild assumptions, we derive the leading error term of order  $O(1/T)$ .

under a fixed- $b$  framework. The explicit form of the leading error term in the approximation provides a clear theoretical explanation for the empirical findings in the literature regarding the direction and magnitude of size distortion for time series with various degrees of dependence. To the best of our knowledge, this is the first time that the leading error terms are made explicit through the higher-order Edgeworth expansion under the fixed-smoothing asymptotics.

Third, we propose a novel bootstrap method for time series, the Gaussian dependent bootstrap, which is able to mimic the second order properties of the original time series and produce a Gaussian bootstrap sample. For the Gaussian location model, we show that the inference based on the Gaussian dependent bootstrap is more accurate than the first order approximation under the fixed-smoothing asymptotics. This seems to be the first time a bootstrap method is shown to be second order correct under the fixed-smoothing asymptotics; see Gonçalves and Vogelsang (2011) for a recent attempt for the moving block bootstrap in the non-Gaussian setting. Fourth, we provide some simulation results that clearly demonstrate the effectiveness of Gaussian dependent bootstrap and the relevance of our higher order theory.

We now introduce some notation. For a vector  $x = (x_1, x_2, \dots, x_{q_0}) \in \mathbb{R}^{q_0}$ , we let  $\|x\| = (\sum_{i=1}^{q_0} x_i^2)^{1/2}$  be the Euclidean norm. For a matrix  $A = (a_{ij})_{i,j=1}^{q_0} \in \mathbb{R}^{q_0 \times q_0}$ , denote by  $\|A\|_2 = \sup_{\|x\|=1} \|Ax\|$  the spectral norm and  $\|A\|_\infty = \max_{1 \leq i, j \leq q_0} |a_{ij}|$  the max norm. Denote by  $[a]$  the integer part of a real number  $a$ . Let  $L^2[0, 1]$  be the space of square integrable functions on  $[0, 1]$ . Denote by  $D[0, 1]$  the space of functions on  $[0, 1]$  which are right continuous and have left limits, endowed with the Skorokhod topology [see Billingsley (1999)]. Denote by “ $\Rightarrow$ ” weak convergence in the  $\mathbb{R}^{q_0}$ -valued function space  $D^{q_0}[0, 1]$ , where  $q_0 \in \mathbb{N}$ . Denote by “ $\rightarrow^d$ ” and “ $\rightarrow^p$ ” convergence in distribution and convergence in probability, respectively. The notation  $N(\mu, \Sigma)$  is used to denote the multivariate normal distribution with mean  $\mu$  and covariance  $\Sigma$ . Let  $\chi_k^2$  be a random variable following  $\chi^2$  distribution with  $k$  degrees of freedom and  $G_k$  be the corresponding distribution function.

The layout of the paper is as follows. Section 2 describes the GMM framework and some high level assumptions. Section 3 presents a general class of estimators for estimating the asymptotic covariance matrix of the GMM estimator. We study the first order fixed-smoothing asymptotics in Section 4. Section 5 contains the higher order expansions of the finite sample distributions of the subsampling  $t$ -statistic and the Wald statistic with the HAC estimator. We introduce the Gaussian dependent bootstrap and the results about its second order accuracy in Section 6. Section 7 concludes. Technical details are gathered in the appendix.

## 2 Basic setup and assumptions

In linear and nonlinear models with moment conditions, it is standard to employ GMM [Hansen (1982)] to estimate the model parameters. We follow the GMM setup as described in KV (2005). Consider a  $d \times 1$  vector of parameters  $\theta \in \Theta \subseteq \mathbb{R}^d$  of interest, where  $\Theta$  is the parameter space. Denote  $\theta_0$  the true parameter of  $\theta$  which is an interior point of  $\Theta$ . Let  $y_t$

denote a vector of observed data and assume the moment conditions

$$E[f(y_t, \theta)] = 0, \quad t = 1, 2, \dots, T \quad (1)$$

hold if and only if  $\theta = \theta_0$ , where  $f(\cdot)$  is  $m \times 1$  vector of functions with  $m \geq d$  and  $\text{rank}(E[\partial f(y_t, \theta_0)/\partial \theta']) = d$ . When  $m > d$ , the parameter  $\theta$  is over-identified with the degree of over-identification  $v = m - d$ . Define the partial sum  $g_t(\theta) = T^{-1} \sum_{j=1}^t f(y_j, \theta)$ . Then the GMM estimator of  $\theta_0$  is given by

$$\hat{\theta}_T = \text{argmin}_{\theta \in \Theta} g_T(\theta)' W_T g_T(\theta), \quad (2)$$

where  $W_T$  is a  $m \times m$  semi-positive definite weighting matrix. Further define

$$G_t(\theta) = \frac{\partial g_t(\theta)}{\partial \theta'} = \frac{1}{T} \sum_{j=1}^t \frac{\partial f(y_j, \theta)}{\partial \theta'}.$$

Using the mean value theorem, we have  $g_T(\hat{\theta}_T) = g_T(\theta_0) + G_T(\tilde{\theta}_T)(\hat{\theta}_T - \theta_0)$ , where  $\tilde{\theta}_T$  is a value between  $\theta_0$  and  $\hat{\theta}_T$ . Note that  $G_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T) = 0$  by the first order condition, which implies that

$$G_T(\hat{\theta}_T)' W_T g_T(\theta_0) + G_T(\hat{\theta}_T)' W_T G_T(\tilde{\theta}_T)(\hat{\theta}_T - \theta_0) = G_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T) = 0.$$

Solving the above equation, we have

$$T^{1/2}(\hat{\theta}_T - \theta_0) = -(G_T(\hat{\theta}_T)' W_T G_T(\tilde{\theta}_T))^{-1} G_T(\hat{\theta}_T)' W_T (T^{1/2} g_T(\theta_0)).$$

To derive the asymptotic distribution of  $\hat{\theta}_T$ , we make the following high-level assumptions as KV (2005) and Sun (2010).

**ASSUMPTION 2.1.**  $\hat{\theta}_T \rightarrow^p \theta_0$ .

**ASSUMPTION 2.2.**  $T^{1/2} g_{\lfloor Tr \rfloor}(\theta_0) \Rightarrow \Delta W_m(r)$  where

$$\Delta \Delta' = \Omega = \sum_{j=-\infty}^{+\infty} E[f(y_t, \theta_0) f(y_{t-j}, \theta_0)'],$$

and  $W_m(r)$  is a  $m$ -dimensional vector of independent standard Brownian motions.

**ASSUMPTION 2.3.**  $G_T(\tilde{\theta}_T) \rightarrow^p G_0$  uniformly for all  $\tilde{\theta}_T$  between  $\hat{\theta}_T$  and  $\theta_0$  where  $G_0 = E[\partial f(y_j, \theta_0)/\partial \theta']$ .

**ASSUMPTION 2.4.** The weighting matrix  $W_T$  is symmetric and semi-positive definite such that  $W_T \rightarrow^p W_0$  and  $G_0' W_0 G_0$  is positive definite.

Under Assumptions 2.1-2.4, it is easy to see that

$$T^{1/2}(\hat{\theta}_T - \theta_0) \rightarrow^d -(G_0' W_0 G_0)^{-1} G_0' W_0 \Delta W_m(1) =^d N(0, V_0),$$

where “ $=^d$ ” denotes “equal in distribution” and the asymptotic covariance matrix  $V_0 := (G'_0 W_0 G_0)^{-1} G'_0 W_0 \Omega W_0 G_0 (G'_0 W_0 G_0)^{-1}$ . To make inference on  $\theta_0$ , we have to estimate  $G_0$ ,  $W_0$  and the LRV matrix  $\Omega$ . Under the above assumptions,  $G_0$  and  $W_0$  can be consistently estimated by their sample counterparts  $G_T(\hat{\theta}_T)$  and  $W_T$  respectively. It remains to estimate the LRV matrix  $\Omega$ . In the next section, we introduce a general class of estimators for  $\Omega$  and  $V_0$ .

### 3 LRV estimators

To present the idea, we focus on the hypothesis testing problem that  $H_{1,0} : r(\theta_0) = 0$  versus the alternative that  $H_{1,a} : r(\theta_0) \neq 0$ , where  $r(\theta)$  is a  $p \times 1$  continuously differentiable function with the first order derivative matrix  $R(\theta) = \partial r(\theta) / \partial \theta'$  and  $p \leq d$ . Let

$$\hat{V}_T = (G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T))^{-1} (G_T(\hat{\theta}_T)' W_T \hat{\Omega}_T W_T G_T(\hat{\theta}_T)) (G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T))^{-1},$$

be an estimator of  $V_0$ , where  $\hat{\Omega}_T$  is the LRV estimate of  $\Omega$ . The Wald statistic for testing  $H_{1,0}$  against  $H_{1,a}$  is defined as

$$F_T = Tr(\hat{\theta}_T)' \hat{D}_T^{-1} r(\hat{\theta}_T) / p, \quad (3)$$

where  $\hat{D}_T = R(\hat{\theta}_T) \hat{V}_T R(\hat{\theta}_T)'$ . The widely used lag window type LRV estimator is given by

$$\hat{\Omega}_T = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \mathcal{K}\left(\frac{i-j}{bT}\right) f(y_i, \hat{\theta}_T) f(y_j, \hat{\theta}_T)', \quad (4)$$

where  $\mathcal{K}(\cdot)$  is a kernel function and  $b$  is the proportion of the truncation lag to the sample size. By setting

$$\hat{u}_i = R(\hat{\theta}_T) (G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T))^{-1} G_T(\hat{\theta}_T)' W_T f(y_i, \hat{\theta}_T),$$

we have

$$\hat{D}_T = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \mathcal{K}\left(\frac{i-j}{bT}\right) \hat{u}_i \hat{u}_j'.$$

When  $\mathcal{K}(\cdot)$  is semi-positive definite<sup>1</sup>, by Mercer's theorem, we have the spectral decomposition,

$$\mathcal{K}(r-t) = \sum_{j=1}^{+\infty} \lambda_j \phi_j(r) \phi_j(t), \quad 0 \leq r, t \leq 1/b, \quad (5)$$

---

<sup>1</sup>A bivariate function  $g(r, s) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  is called semi-positive definite if for any positive integer  $n$ , we have  $\sum_{i,j=1}^n c_i c_j g(a_i, a_j) \geq 0$  for all  $(a_1, a_2, \dots, a_n)$  and  $(c_1, c_2, \dots, c_n)$  in  $\mathbb{R}^n$ . Here we assume that  $\mathcal{K}(r-s) = g(r, s)$  is semi-positive definite.

where  $\{\lambda_j\}$  and  $\{\phi_j\}$  are the eigenvalues and orthonormal eigenfunctions corresponding to the kernel function respectively. We thus have the representation,

$$\hat{D}_T = \sum_{s=1}^K \lambda_s \left\{ \frac{1}{\sqrt{T}} \sum_{i=1}^T \phi_s \left( \frac{i}{bT} \right) \hat{u}_i \right\} \left\{ \frac{1}{\sqrt{T}} \sum_{j=1}^T \phi_s \left( \frac{j}{bT} \right) \hat{u}'_j \right\},$$

with  $K = +\infty$ . In the traditional asymptotics,  $b$  goes to zero as  $T$  increases which is referred as the small- $b$  asymptotics. When  $b \in (0, 1]$  is held fixed, it corresponds to the fixed- $b$  asymptotics in KV (2005). As pointed out in some recent studies [see e.g., Bester et al. (2009); Sun (2011a; 2011b); Chen and Qu (2010)],  $K$  can also be held as a fixed positive integer, which can lead to a more accurate first order approximation. In light of these recent findings, we introduce a general class of estimators to estimate the LRV matrix. With a slight abuse of notation, we let  $\{\phi_s(t)\}_{s=1}^K$  be a sequence of linearly independent functions<sup>2</sup> in  $L^2[0, 1/b]$  and  $\{\lambda_j\}$  be a sequence of nonnegative weights such that  $\sum_{j=1}^K \lambda_j = 1$ . Note that  $\lambda_j$ 's in (5) are nonnegative when we consider semi-positive definite kernels in (4). Further let  $V_s = \frac{1}{\sqrt{T}} \sum_{i=1}^T \phi_s \left( \frac{i}{bT} \right) \hat{u}_i$ , be the normalized inner product between  $\{\hat{u}_i\}_{i=1}^T$  and  $\{\phi_s(i/(bT))\}_{i=1}^T$ . Define  $R = (R_{ij})_{i,j=1}^K$  with  $R_{ij} = \int_0^1 \tilde{\phi}_i(t/b) \tilde{\phi}_j(t/b) dt$ , where  $\tilde{\phi}_s(t/b) = \phi_s(t/b) - \int_0^1 \phi_s(t/b) dt$ , and  $L = (L_{ij})_{i,j=1}^K$  an upper triangular matrix based on the Cholesky decomposition of  $R^{-1}$ , i.e.,  $L'L = R^{-1}$ . Define  $V = (V'_1, V'_2, \dots, V'_K)'$  and

$$V^* = (V_1^{*'}, V_2^{*'}, \dots, V_K^{*'})' = (L \otimes I_p)V,$$

where  $V_i^* = \sum_{j=1}^K L_{ij} V_j$  for  $1 \leq i \leq K$ . Then the general LRV estimator is given by

$$\hat{D}_T = \sum_{s=1}^K \lambda_s V_s^* V_s^{*'} = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \left\{ \sum_{s=1}^K \lambda_s \sum_{m=1}^K L_{sm} \phi_m \left( \frac{i}{bT} \right) \sum_{l=1}^K L_{sl} \phi_l \left( \frac{j}{bT} \right) \right\} \hat{u}_i \hat{u}'_j, \quad (6)$$

and the test statistic based on the general LRV estimator is defined as,

$$F_T = [\sqrt{T}r(\hat{\theta}_T)]' \hat{D}_T^{-1} [\sqrt{T}r(\hat{\theta}_T)]/p. \quad (7)$$

The matrix  $R$  is introduced for orthogonalization so that the limiting distribution of the test statistic  $F_T$  does not depend on the basis functions. Note that the choice of  $R$  is not unique (See Example 3.3). In what follows, we shall show that the recently developed nonparametric series covariance estimator [Sun (2011a; 2011b)], the recursive subsampling-based covariance estimator [Chen and Qu (2010)] and the cluster covariance estimator (CCE) [Bester et al. (2009)] are all special cases of the general LRV estimator. Throughout Examples 3.1-3.3, we set  $b = 1$  and  $\lambda_j = 1/K$  for  $j = 1, 2, \dots, K$ .

**EXAMPLE 3.1.** Let  $\{\phi_s(t)\}_{s=1}^K$  be a sequence of orthonormal basis functions with  $\int_0^1 \phi_s(t) dt = 0$ .

---

<sup>2</sup>A set of elements  $\{\psi_i\}_{i=1}^K$  in a real valued vector space is called linearly independent if and only if  $\sum_{i=1}^K a_i \psi_i = \mathbf{0} \Rightarrow a_i = 0$  for  $i = 1, 2, \dots, K$ . Here  $\mathbf{0}$  denotes the null element in the vector space.



Then we have  $R = I_{K \times K}$  and  $\hat{D}_T = \frac{1}{K} \sum_{j=1}^K V_j V_j'$ , where  $V_j = \frac{1}{\sqrt{T}} \sum_{i=1}^T \phi_j(i/T) \hat{u}_i$ . When  $\phi_s(t) = \sqrt{2} \sin(2\pi st)$  (or  $\phi_s(t) = \sqrt{2} \cos(2\pi st)$ ),  $s = 1, 2, \dots, K$ , it is straightforward to see that the LRV estimator corresponds to the series estimator considered in Sun (2011a; 2011b). In this case, the LRV estimator involves projecting the data onto a set of orthonormal basis and using the sample variance of the projection vectors, namely  $\hat{D}_T$ .

**EXAMPLE 3.2.** For any fixed  $K$  with  $K \leq T$ , we consider the basis function  $\phi_s(t) = \mathbf{I}\{0 < t \leq s/(K+1)\}$ ,  $s = 1, 2, \dots, K$ , where  $\mathbf{I}$  denotes the indicator function. Simple calculation gives us  $R_{ij} = \int_0^1 \tilde{\phi}_i(t) \tilde{\phi}_j(t) dt = \min(i, j)/(K+1) - (ij)/(K+1)^2$ ,<sup>3</sup> and  $\hat{D}_T = \frac{1}{K} \sum_{s=1}^K V_s^* V_s^{*'}$ , where

$$V_s^* = \sqrt{\frac{K+1}{T}} \left( \sqrt{\frac{s+1}{s}} \sum_{i=1}^{\lfloor \frac{Ts}{K+1} \rfloor} \hat{u}_i - \sqrt{\frac{s}{s+1}} \sum_{i=1}^{\lfloor \frac{T(s+1)}{K+1} \rfloor} \hat{u}_i \right),$$

with  $s = 1, 2, \dots, K$  and  $V_{K+1} = 0$ . Therefore, the general LRV estimator reduces to the recursive subsampling-based estimator in Chen and Qu (2010), where the idea is to divide the full sample into  $K+1$  recursive subsamples and construct a normalization matrix based on the subsamples.

**EXAMPLE 3.3.** Let  $\{A_j\}_{j=1}^K$  be a partition of the unit intervals  $[0, 1]$  with  $K > p$ . Suppose  $A_j$  is a finite union of disjoint intervals in  $[0, 1]$ . Let  $\phi_s(t) = \mathbf{I}(t \in A_s)$ ,  $s = 1, 2, \dots, K$ . If we set  $R_{ij} = \int_0^1 \phi_i(t) \phi_j(t) dt$ , then  $L = \text{diag}(1/\sqrt{|A_1|}, 1/\sqrt{|A_2|}, \dots, 1/\sqrt{|A_K|})$ , where  $|A|$  denotes the Lebesgue measure of the set  $A$ . Further assume  $|A_1| = |A_2| = \dots = |A_K| = 1/K$ , then we have

$$\hat{D}_T = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \sum_{s=1}^K \mathbf{I}(i/T \in A_s) \mathbf{I}(j/T \in A_s) \hat{u}_i \hat{u}_j' = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \mathbf{I}(i, j \in \text{the same group}) \hat{u}_i \hat{u}_j',$$

where  $i$  is in group  $s$  if and only if  $i/T \in A_s$ ,  $s = 1, 2, \dots, K$ . In this case, the general LRV estimator is the same as the CCE considered in Bester et al. (2009), where the idea is to utilize the group structure in the observations and construct a covariance estimator based on the parameter estimates in each group. Using similar arguments in Sun (2010), we can show that

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor Tr \rfloor} \hat{u}_i \Rightarrow \Lambda B_p(r),$$

where  $\Lambda$  is an invertible matrix such that

$$\Lambda \Lambda' = R(\theta_0) (G_0' W_0 G_0)^{-1} G_0' W_0 \Omega W_0 G_0 (G_0' W_0 G_0)^{-1} R'(\theta_0)$$

and  $B_p(r)$  denotes a  $p$ -dimensional vector of independent Brownian bridges. It implies that

$$\frac{1}{\sqrt{T}} \sum_{i \in \text{sth group}} \hat{u}_i \rightarrow^d \Lambda \int_{A_s} dB_p(r) =^d \frac{1}{\sqrt{K}} \Lambda (Z_s - \bar{Z}),$$

---

<sup>3</sup>In this case, we have  $L_{ii} = \sqrt{\frac{(K+1)(i+1)}{i}}$ ,  $L_{i,i+1} = -\sqrt{\frac{(K+1)i}{i+1}}$  and  $L_{ij} = 0$  otherwise.



and

$$\hat{D}_T \rightarrow^d \frac{1}{K} \Lambda \sum_{s=1}^K (Z_s - \bar{Z})(Z_s - \bar{Z})' \Lambda',$$

where  $(Z'_1, Z'_2, \dots, Z'_K)' \sim N(0, I_K \otimes I_p)$  and  $\bar{Z} = \sum_{s=1}^K Z_s / K$ . When  $p = 1$ , it is well known that

$$\sum_{s=1}^K (Z_s - \bar{Z})^2 \stackrel{d}{=} \chi_{K-1}^2,$$

which implies  $\sqrt{F_T} \rightarrow^d \sqrt{\frac{K}{K-1}} |t_{K-1}|$  under the  $H_{1,0}$ . Note that  $\sqrt{\frac{K-1}{K}} F_T$  coincides with the subsampling-based  $t$ -statistic in Ibragimov and Müller (2010) when we consider a location model and  $r(\theta_0) = \theta_0 - \theta^*$  for a specific value  $\theta^*$ . When  $p > 1$ , we have  $F_T \rightarrow^d \frac{K}{K-p} F_{p, K-p}$ . It is worth noting that the choice of  $R = (R_{ij})$  with  $R_{ij} = \int_0^1 \tilde{\phi}_i(t) \tilde{\phi}_j(t) dt$  is also valid. In this case, the limiting distribution of  $F_T$  would be a scaled  $F$  distribution with  $p$  numerator and  $K - p + 1$  denominator degrees of freedom [see Theorem 4.1].

**REMARK 3.1.** For the subsampling-based inference, Assumption 2.2 can be relaxed by the assumptions which guarantee the finite dimensional convergence of  $\left( \frac{1}{\sqrt{|\mathcal{G}_1|}} \sum_{i \in \mathcal{G}_1} \hat{u}_i, \dots, \frac{1}{\sqrt{|\mathcal{G}_K|}} \sum_{i \in \mathcal{G}_K} \hat{u}_i \right)$ . Here  $\mathcal{G}_i$  is the set index for the  $i$ th group and  $|\cdot|$  denotes the cardinality. When heteroscedasticity is present across different groups, the  $t$ -statistic tends to be conservative [see Ibragimov and Müller (2010)].

## 4 First order fixed-smoothing asymptotics

In what follows, we consider the first order fixed-smoothing asymptotics of the test statistic  $F_T$  based on the general LRV estimator under the null hypothesis and local alternatives. To emphasize the dependence on the smoothing parameter  $K$ , we shall use the notation  $F_T(K)$  instead of  $F_T$ .

**THEOREM 4.1.** Suppose  $p \leq K < \infty$  and  $b \in (0, 1]$  are both fixed. Let  $R = (R_{ij})_{i,j=1}^K$  with  $R_{ij} = \int_0^1 \tilde{\phi}_i(t/b) \tilde{\phi}_j(t/b) dt$  in the general LRV estimator. Further assume that  $\phi_j(t)$  is continuously differentiable almost everywhere for  $j = 1, 2, \dots, K$ . Under Assumptions 2.1-2.4 and  $H_{1,0}$ , we have

$$F_T(K) \rightarrow^d Q_{p,K} := U_p' D_p^{-1} U_p / p, \quad (8)$$

where  $D_p = \sum_{j=1}^K \lambda_j \eta_j \eta_j'$ ,  $\{\eta_j\}_{j=1}^K$  and  $U_p$  are independent and identically distributed (iid) as  $N(0, I_p)$ . In particular, if  $\lambda_j = 1/K$  for  $j = 1, 2, \dots, K$ , we get

$$F_T(K) \rightarrow^d \frac{K}{K - p + 1} F_{p, K-p+1}. \quad (9)$$

**REMARK 4.1.** When the weights  $\lambda_j$ 's are not equal and  $p = 1$ ,  $D_p$  is a weighted sum of independent  $\chi_1^2$  random variables. The limiting null distribution  $Q_{p,K}$  can be further approximated by a scaled  $F$  distribution with the parameters chosen properly to match the first two moments

[see Sun (2010)]. Compared to Sun (2011a), we do not make the assumption that  $\int_0^1 \phi_i(t)dt = 0$  and we allow the basis functions to be non-orthonormal (see Example 3.2). It is also worth noting that the above results hold when  $\phi_s(t) = \mathbf{I}(t \in A_s)$  with  $A_s$  being a finite union of disjoint intervals in  $[0, 1]$ .

**THEOREM 4.2.** *Consider the local alternatives  $H'_{1,a} : r(\theta_0) = c/\sqrt{T}$  with  $c \neq \mathbf{0} \in \mathbb{R}^p$ . Under the same assumptions in Theorem 4.1 with  $\lambda_j = 1/K$ , we have*

$$F_T(K) \rightarrow^d \frac{K}{K-p+1} F_{p,K-p+1,c'(R(\theta_0)V_0R(\theta_0)')^{-1}c},$$

where  $F_{a,b,\delta}$  denotes the noncentral  $F$  distribution with degrees of freedom  $a$  and  $b$ , and noncentral parameter  $\delta$ .

The theorem shows that the test  $F_T(K)$  has non-trivial power against the local alternatives of order  $1/\sqrt{T}$  and it is seen to be consistent if  $\|c\| \rightarrow +\infty$  as  $T \rightarrow +\infty$ . When  $b$  is fixed and  $K$  satisfies  $1/K + K/T \rightarrow 0$ , we can show that the general LRV estimator is consistent and the limiting distribution of  $F_T(K)$  is  $\chi_p^2/p$ . Since the main focus of this article is on the fixed-smoothing asymptotics (i.e.,  $K$  is fixed), we do not present the proof but would expect the argument to be similar to Sun (2011a).

## 5 Higher order expansions

This paper is partially motivated by recent studies on the ERP for the Gaussian location model by Jansson (2004) and Sun et al. (2008), who showed that the ERP is of order  $O(1/T)$  under the fixed- $b$  asymptotics, which is smaller than the ERP under the small- $b$  asymptotics. A natural question is to what extent the ERP result can be extended to the above-mentioned methods in Section 1 under the fixed-smoothing asymptotics. Following Jansson (2004) and Sun et al. (2008), we focus on the inference of the mean of a univariate Gaussian stationary time series or equivalently a Gaussian location model. We expect that the higher order terms in the asymptotic expansion under the Gaussian assumption will also show up in the general expansion without the Gaussian assumption.

### 5.1 Expansion for the finite sample distribution of subsampling-based $t$ -statistic

In this part, we investigate the Edgeworth expansion of the distribution of the subsampling-based  $t$ -statistic [Ibragimov and Müller (2010)]. Here we treat the subsampling-based  $t$ -statistic and other cases separately, because the  $t$ -statistic corresponds to a different choice of orthogonalization matrix  $R$  as explained in Example 3.3. Given the observations  $\mathcal{X} = (X_1, X_2, \dots, X_T)$ , we divide the sample into  $K$  approximately equal sized groups of consecutive observations. The observation  $X_i$  is in the  $j$ -th group if and only if  $i \in \mathcal{G}_j = \{s \in \mathbb{Z} : (j-1)T/K < s \leq$

$jT/K\}$ ,  $j = 1, 2, \dots, K$ . Assume that the time series  $\{X_i\}$  is stationary and Gaussian with mean  $\mu$  and autocovariance function  $\gamma_X(i-j) = E[(X_i - \mu)(X_j - \mu)]$ . Define the sample mean of the  $k$ -th group as

$$\hat{\mu}_k = \frac{1}{|\mathcal{G}_k|} \sum_{i \in \mathcal{G}_k} X_i, \quad k = 1, 2, \dots, K.$$

Let  $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_K)'$ ,  $\bar{\mu}_n = \frac{1}{K} \sum_{i=1}^K \hat{\mu}_i$  and  $S_n^2 = \frac{1}{K-1} \sum_{i=1}^K (\hat{\mu}_i - \bar{\mu}_n)^2$ . Then the subsampling-based  $t$ -statistic for testing the null hypothesis  $H_{2,0} : \mu = \mu_0$  versus the alternative  $H_{2,a} : \mu \neq \mu_0$ , is given by

$$T_K = \frac{\sqrt{K}(\bar{\mu}_n - \mu_0)}{S_n} = \frac{\sqrt{K}(\bar{\mu}_n - \mu_0)}{\left\{ \frac{1}{K-1} \sum_{i=1}^K (\hat{\mu}_i - \bar{\mu}_n)^2 \right\}^{1/2}}. \quad (10)$$

Our goal here is to develop an Edgeworth expansion of  $P(|T_K| \leq x)$  when  $K$  is fixed and sample size  $T \rightarrow \infty$ . Denote by  $t_k$  a random variable following  $t$  distribution with  $k$  degrees of freedom. The following theorem gives the higher order expansion under the Gaussian assumption.

**THEOREM 5.1.** *Assume that  $\{X_t\}$  is a stationary Gaussian time series with autocovariance function  $\{\gamma_X(h)\}$  satisfying that  $\sigma^2 := \sum_{h=-\infty}^{+\infty} \gamma_X(h) > 0$  and  $\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty$ . Further suppose that  $|\mathcal{G}_1| = |\mathcal{G}_2| = \dots = |\mathcal{G}_K|$  and  $K$  is fixed. Then under  $H_{2,0}$ , we have*

$$\sup_{x \in [0, +\infty)} |P(|T_K| \leq x) - \Psi_K(x)| = O(1/T^2), \quad (11)$$

where  $\Psi_K(x) = P(|t_{K-1}| \leq x) - \frac{B}{2\sigma^2 T} \Upsilon(x, K)$  with

$$\Upsilon(x, K) = -K^2 P(|t_{K-1}| \leq x) + (K+1) E \left[ \chi_{K-1}^2 G_1 \left( \frac{\chi_{K-1}^2 x^2}{K-1} \right) \right] - E \left[ \chi_1^2 G_{K-1} \left( \frac{(K-1)\chi_1^2}{x^2} \right) \right] + 1,$$

and  $B = \sum_{h=-\infty}^{+\infty} |h| \gamma_X(h)$ .

From the above expression, we see that the leading error term is of order  $O(1/T)$  and the magnitude and direction of the error depend upon  $B/\sigma^2$ , which is related to the second order properties of time series, and  $\Upsilon(x, K)$ , which is independent of the dependence structure of  $\{X_t\}$  and can be approximated numerically for given  $x$  and  $K$ . Figure 1 plots the approximated values of  $\Upsilon(t_{K-1}(1-\alpha), K)/K$  for different  $K$  and  $\alpha$ , where  $t_{K-1}(1-\alpha)$  denotes the  $100(1-\alpha)\%$  quantile of the  $t$  distribution with  $K-1$  degrees of freedom. It can be seen from Figure 1 that  $\Upsilon(t_{K-1}(1-\alpha), K)/K$  increases rapidly for  $K < 10$  and it becomes stable for relatively large  $K$ . For each  $K \geq 2$ ,  $\Upsilon(t_{K-1}(1-\alpha), K)/K$  is an increasing function of  $\alpha$ . In the simulation work of Ibragimov and Müller (2010) (see Figure 2 therein), they found that the size of the subsampling based  $t$ -test is relatively robust to the correlations if  $K$  is small (say  $K = 4$  in their simulation). This finding is in fact supported by our theory. For  $K \leq 4$ , the magnitude of  $\Upsilon(x, K)$  is rather small, so the leading error term is small across a range of correlations. As  $K$  increases, the first order approximation deteriorates, which is reflected in the increasing

magnitude of  $\Upsilon(t_{K-1}(1 - \alpha), K)$  with respect to  $K$ .

Notice that  $\Upsilon(t_{K-1}(1 - \alpha), K)$  is always positive and  $\sigma^2 > 0$  by assumption, so the sign of the leading error term, i.e.,  $-\frac{B}{2\sigma^2 T}\Upsilon(x, K)$  is determined by  $B$ . When  $B > 0$  (e.g., AR(1) process with positive coefficient), the first order based inference tends to be oversized and conversely it tends to be undersized when  $B < 0$  (e.g., MA(1) process with negative coefficient). Some simulations for AR(1) and MA(1) models in the Gaussian location model support these theoretical findings. We decide not to report these results to conserve space. Given the sample size  $T$ , the size distortion for the first order based inference may be severe if the ratio  $B/\sigma^2$  is large. For example, this is the case for AR(1) model,  $X_t = \rho X_{t-1} + \varepsilon_t$ , as the correlation  $\rho$  gets closer to 1. As indicated by Figure 1, we show in the following proposition that  $\Upsilon(t_{K-1}(1 - \alpha), K)/K$  converges as  $K \rightarrow \infty$ .

**PROPOSITION 5.1.** *As  $K \rightarrow +\infty$ , we have  $\Upsilon(x, K)/K = 2x^2 G'_1(x^2) + O(1/K)$ , for any fixed  $x \in \mathbb{R}$ .*

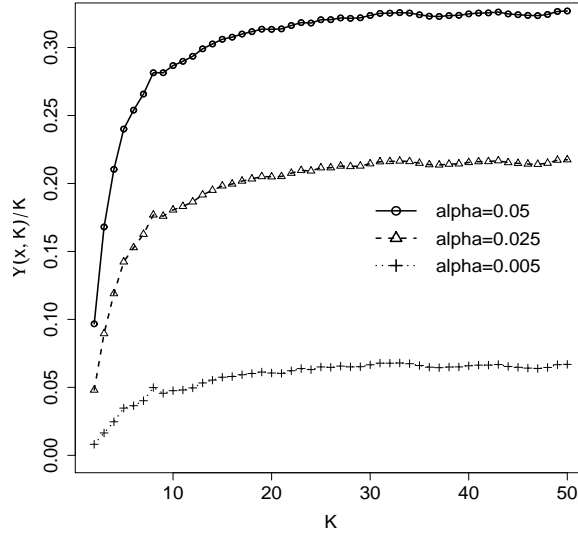


Figure 1: Simulated values of  $\Upsilon(t_{K-1}(1 - \alpha), K)/K$  based on 500,000 replications.

Under the local alternative  $H'_{2,a} : \mu = \mu_0 + (\delta\sigma)/\sqrt{T}$  with  $\delta \neq 0$ , we can derive a similar expansion for  $T_K$  with  $K$  fixed. Formally let  $Z$  be a random variable following the standard normal distribution and  $\mathcal{S}_{K-1} = \sqrt{\chi^2_{K-1}/(K-1)}$  with the  $\chi^2_{K-1}$  distribution being independent with  $Z$ . Then the quantity  $t_{K-1,\delta} = (Z + \delta)/\mathcal{S}_{K-1}$  follows a noncentral  $t$  distribution with noncentral parameter  $\delta$ . Define  $e_1(x) = E[\mathbf{I}\{|t_{K-1,\delta}| > x\}Z^2]$  and  $e_2(x) = E[\mathbf{I}\{|t_{K-1,\delta}| > x\}\chi^2_{K-1}]$ . Then under the local alternative, we have

$$P(|T_K| \leq x) = P(|t_{K-1,\delta}| \leq x) - \frac{B}{2\sigma^2 T}\Upsilon_\delta(x, K) + O(1/T^2),$$

where  $\Upsilon_\delta(x, K) = K^2 P(|t_{K-1,\delta}| > x) - e_1(x) - (K+1)e_2(x)$ . For fixed  $\delta$ ,  $P(|t_{K-1,\delta}| > t_{K-1}(1 - \alpha))$  is a monotonic increasing functions of  $K$ . Unreported numerical study shows

that  $\Upsilon_\delta(t_{K-1}(1-\alpha), K)$  is roughly monotonic with respect to  $K$  for  $\delta \in (0, 4]$ , which suggests that larger  $K$  tends to deliver more power when  $B > 0$ . Combined with the previous discussion, we see that the choice of  $K$  leads to a trade-off between the size distortion and power loss.

**REMARK 5.1.** Theorem 5.1 gives the ERP and the exact form of the leading error term under the fixed- $K$  asymptotics. The high order expansion derived here is based on an expansion of the density function of  $(\hat{\mu}_1, \dots, \hat{\mu}_K)$  which is made possible by the Gaussian assumption. Expansion for a distribution function or equivalently characteristic function has been used in the high order expansion of the finite sample distribution under the Gaussian assumption [see e.g., Velasco and Robinson (2001), and Sun et al. (2008)]. With  $K$  fixed in the asymptotics, the variance of the LRV estimator is captured by the first order fixed- $K$  limiting distribution and the bias of the LRV estimator is reflected in the higher order term  $-\frac{B}{2\sigma^2 T}\Upsilon(x, K)$ .

**REMARK 5.2.** When the number of groups  $K$  grows slowly with the sample size  $T$ , the Edgeworth expansion for  $T_K$  was developed for  $P(T_K \leq x)$  in Lahiri (2007; 2010) under the general non-Gaussian setup. The expansion given here is different from the usual Edgeworth expansion under the increasing-smoothing asymptotics in terms of the form and the convergence rate. Using the same argument, we can show that under the fixed- $K$  asymptotics, the leading error term in the expansion of  $P(T_K \leq x)$  is of order  $O(1/T)$  under the Gaussian assumption. In the non-Gaussian case, we conjecture that the order of the leading error term is  $O(1/\sqrt{T})$ , which is due to the effect of the third and fourth order cumulants.

The higher order Edgeworth expansion results in Sun et al. (2008) suggest that the fixed- $b$  based approximation is a refinement of the approximation provided by the limiting distribution derived under the small- $b$  asymptotics. In a similar spirit, it is natural to ask if the fixed- $K$  based approximation refines the first order approximation under the increasing- $K$  asymptotics. To address this question, we consider the expansion under the increasing-smoothing asymptotics, where  $K$  grows slowly with the sample size  $T$ .

**PROPOSITION 5.2.** *Under the same conditions in Theorem 5.1 but with  $\lim_{T \rightarrow \infty} (1/K + K/T) = 0$ , we have*

$$P(|T_K| \leq x) = G_1(x^2) + \frac{1}{K-1}x^4 G_1''(x^2) - \frac{BK}{T\sigma^2}x^2 G_1'(x^2) + O(1/T). \quad (12)$$

**REMARK 5.3.** Since

$$P(|t_{K-1}| \leq x) = G_1(x^2) + \frac{1}{K-1}x^4 G_1''(x^2) + O(1/K^2)$$

[see e.g., Sun(2011a)], we know that the fixed- $K$  based approximation captures the first two terms in (12), whereas the increasing- $K$  based approximation (i.e.,  $\chi_1^2$ ) only captures the first term. In view of Proposition 5.1, it is not hard to see that

$$\Psi_K(x) = G_1(x^2) + \frac{1}{K-1}x^4 G_1''(x^2) - \frac{BK}{T\sigma^2}x^2 G_1'(x^2) + O(1/K^2) + O(1/T),$$

which implies that the fixed- $K$  based expansion is able to capture all the three error terms in (12) as the smoothing parameter  $K \rightarrow \infty$  with  $T^{1/3} = o(K)$ . Loosely speaking, this suggests that the fixed- $K$  based expansion holds for a broad range of  $K$  and it gets close to the corresponding increasing- $K$  based expansion when  $K$  is large.

## 5.2 Fixed- $b$ expansion (with $K = +\infty$ )

Consider a semi-positive definite bivariate kernel  $\mathcal{G}(\cdot, \cdot)$  which satisfies the spectral decomposition

$$\mathcal{G}(r, t) = \sum_{j=1}^{+\infty} \lambda_j \phi_j(r) \phi_j(t), \quad 0 \leq r, t \leq 1, \quad (13)$$

where  $\{\phi_j\}$  are the eigenfunctions and  $\{\lambda_j\}$  are the eigenvalues which are in a descending order, i.e.,  $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ . Here we set  $b = 1$  for the convenience of presentation. See Remark 5.4 for the case  $b \in (0, 1)$ . Define the projection vectors  $\xi_j = \frac{1}{\sqrt{T}} \sum_{i=1}^T \phi_j^0(i/T) X_i$  with  $\phi_j^0(t) = \phi_j(t) - \frac{1}{T} \sum_{i=1}^T \phi_j(i/T)$  for  $j = 1, 2, \dots + \infty$ . Here the dependence of  $\xi_j$  on  $T$  is suppressed to simplify the notation. Following Sun (2011a), we limit our attention to the case  $\int_0^1 \phi_j(t) dt = 0$  (e.g., fourier basis and Haar wavelet basis)<sup>4</sup>. Then the LRV estimator can be written as

$$\hat{D}_T = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \mathcal{G}\left(\frac{i}{T}, \frac{j}{T}\right) (X_i - \bar{X}_T)(X_j - \bar{X}_T) = \sum_{i=1}^{+\infty} \lambda_i \xi_i^2,$$

where  $\bar{X}_T$  is the sample mean. Again we focus on the hypothesis testing problem ( $H_{2,0}$  versus  $H_{2,a}$ ). Define a sequence of random variables

$$F_T(J) = \frac{\xi_0^2}{\sum_{j=1}^J \lambda_j \xi_j^2}, \quad J = 1, 2, \dots, \infty$$

with  $\xi_0 = \frac{1}{\sqrt{T}} \sum_{i=1}^T (X_i - \mu_0)$ . Our test statistic is  $F_T(\infty) = \xi_0^2 / \hat{D}_T$ . Let  $\{v_i\}_{i=0}^{+\infty}$  be a sequence of iid standard normal random variables. Further define  $\mathcal{F}_\infty(v) = \frac{v_0^2}{\sum_{j=1}^\infty \lambda_j v_j^2}$  and

$$\psi_T(x) = \frac{1}{2\sigma^2} \sum_{i=0}^{+\infty} (\text{var}(\xi_i) - \sigma^2) E[(v_i^2 - 1) \mathbf{I}\{\mathcal{F}_\infty(v) \leq x\}]. \quad (14)$$

---

<sup>4</sup>Given any semi-positive definite kernel  $\mathcal{G}(\cdot, \cdot)$ , we can define the demeaned kernel,

$$\tilde{\mathcal{G}}(r, t) = \mathcal{G}(r, t) - \int_0^1 \mathcal{G}(s, t) ds - \int_0^1 \mathcal{G}(r, p) dp + \int_0^1 \int_0^1 \mathcal{G}(s, p) ds dp.$$

Suppose  $\tilde{\mathcal{G}}(\cdot, \cdot)$  admits the spectral decomposition  $\tilde{\mathcal{G}}(r, t) = \sum_{i=1}^{+\infty} \tilde{\lambda}_i \tilde{\phi}_i(r) \tilde{\phi}_i(t)$  with  $\{\tilde{\phi}_i\}$  and  $\{\tilde{\lambda}_i\}$  being the eigenfunctions and eigenvalues respectively. Notice that

$$\int_0^1 \int_0^1 \tilde{\mathcal{G}}(r, t) dr dt = \sum_{i=1}^{+\infty} \tilde{\lambda}_i \left( \int_0^1 \tilde{\phi}_i(t) dt \right)^2 = 0,$$

which implies  $\int_0^1 \tilde{\phi}_i(t) dt = 0$  whenever  $\lambda_i > 0$ , i.e., the eigenfunctions of the demeaned kernel  $\tilde{\mathcal{G}}$  are all mean zero.

The following theorem establishes the asymptotic expansion of the distribution of  $F_T(\infty)$ .

**THEOREM 5.2.** *Assume the kernel  $\mathcal{G}(\cdot, \cdot)$  satisfies the following conditions:*

- (1) *Suppose the second derivatives of the eigenfunctions  $\{\phi_i^{(2)}\}_{i=1}^\infty$  exist. Further assume that the eigenfunctions are mean zero and satisfy that  $\sup_{1 \leq i \leq J} \sup_{t \in [0,1]} |\phi_i^{(j)}(t)| < CJ^j$ , for  $j = 0, 1, 2$ ,  $J \in \mathbb{N}$ , and some constant  $C$  which does not depend on  $j$  and  $J$ ;<sup>5</sup>*
- (2) *The eigenvalues  $\lambda_n = O(1/n^a)$ , for some  $a > 19$ .*

*Under the assumption that  $\{X_t\}$  is a stationary Gaussian time series with  $\sigma^2 > 0$  and  $\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty$ , and the null hypothesis  $H_{2,0}$ , we have  $\sup_{x \in [0, +\infty)} |\psi_T(x)| = O(1/T)$  and*

$$\sup_{x \in [0, +\infty)} |P(F_T(\infty) \leq x) - P(\mathcal{F}_\infty(v) \leq x) - \psi_T(x)| = o(1/T). \quad (15)$$

Assume  $\int_0^1 \mathcal{G}(r, r) dr = \sum_{j=1}^{+\infty} \lambda_j = 1$ . As seen from Theorem 5.2, the bias of the LRV estimator (i.e.,  $\sum_{i=1}^\infty \lambda_i (\text{var}(\xi_i) - \sigma^2)$ ) is reflected in the leading error term  $\psi_T(x)$ , which is a weighted sum of the relative difference of  $\text{var}(\xi_i)$  and  $\sigma^2$ . Note that the difference  $\text{var}(\xi_i) - \sigma^2$  relies on the second order properties of the time series and the eigenfunctions of  $\mathcal{G}(\cdot, \cdot)$ , and the weight  $E[(v_i^2 - 1)\mathbf{I}\{\mathcal{F}_\infty(v) \leq x\}]$  which depends on the eigenvalues of  $\mathcal{G}(\cdot, \cdot)$  is of order  $O(\lambda_i)$ , as seen from the arguments used in the proof of Theorem 5.2.

**REMARK 5.4.** In the appendix (see Lemma 8.6), we establish the higher order expansion for the Wald statistic based on the general LRV estimators considered in Section 3. This result can be viewed as a special case of Theorem 5.2 when the kernel function belongs to a finite dimensional space. For  $0 < b \leq 1$ , we define  $\mathcal{G}_b(\cdot, \cdot) = \mathcal{G}(\cdot/b, \cdot/b)$ . If  $\mathcal{G}(\cdot, \cdot)$  is semi-positive definite on  $[0, 1/b]^2$ , it satisfies the spectral decomposition  $\mathcal{G}_b(r, t) = \sum_{j=1}^{+\infty} \lambda_{j,b} \phi_{j,b}(r) \phi_{j,b}(t)$  with  $0 \leq r, t \leq 1/b$ .<sup>6</sup> Our result can then be extended to the case where  $b < 1$  if the assumptions in Theorem 5.2 hold for  $\{\lambda_{j,b}\}$  and  $\{\phi_{j,b}\}$ . It is worth noting that our result is established under different assumptions as compared to Theorem 6 in Sun et al. (2008), where the bivariate kernel is defined as  $\mathcal{G}(r, t) = \mathcal{K}(r - t)$  and the technical assumption  $b < 1/(16 \int_{-\infty}^{+\infty} |\mathcal{K}(r)| dr)$  is required, which rules out the case  $b = 1$  for most kernels. Here we provide an alternative way of proving the  $O(1/T)$  ERP when the eigenfunctions are mean zero. Furthermore, we provide the exact form of the leading error term which has not been obtained in the literature.

In econometric and statistical literature, the bivariate kernel  $\mathcal{G}(\cdot, \cdot)$  is usually defined through a semi-positive definite univariate kernel  $\mathcal{K}(\cdot)$  i.e.,  $\mathcal{G}(r, t) = \mathcal{K}(r - t)$ . In what follows, we make several remarks regarding this special case.

<sup>5</sup>This condition is on the oscillation of the basis functions. It is satisfied by the fourier basis functions  $\{\sqrt{2} \cos(2\pi jt), \sqrt{2} \sin(2\pi jt)\}_{j=1}^\infty$

<sup>6</sup>Given a semi-positive definite kernel  $\mathcal{G}_b(r, t)$ , its eigencomponents can be obtained by solving a homogenous Fredholm integral equation of the second kind, where the solutions can be approximated numerically when analytical solutions are unavailable. When  $\mathcal{G}(r, t) = \mathcal{K}(r - t)$ , it was shown in Knessl and Keller (1991) that under suitable assumptions on  $\mathcal{K}(\cdot)$ ,  $\lambda_{j,b} = b \int_{-\infty}^{+\infty} \mathcal{K}(r) dr - (\pi^2 j^2 b^3 / 2) \int_{-\infty}^{+\infty} r^2 \mathcal{K}(r) dr + o(b^3)$  and  $\phi_{j,b} \approx \sqrt{2} \sin(\pi j x)$  for  $x$  bounded away from 0 and 1 as  $b \rightarrow 0$ , which implies that  $\lambda_{M,b} / \lambda_{1,b} \rightarrow 1$  for any fixed  $M \in \mathbb{N}$  and  $b \rightarrow 0$ .



**REMARK 5.5.** The assumption on the eigenvalues is satisfied by the bivariate kernel defined through the QS kernel and the Daniel kernel with  $0 < b \leq 1$ , and the Tukey-Hanning kernel with  $b = 1$  because these kernels are analytical on the corresponding regions and their eigenvalues decay exponentially fast [see Little and Reade (1984)]. Note that the assumption does not hold for the Bartlett kernel because the decay rate of its eigenvalues is of order  $O(1/n^2)$ . For the demeaned Tukey-Hanning kernel with  $b = 1$ , we have that the eigenfunctions  $\phi_1(t) = \sqrt{2} \cos \pi t$  and  $\phi_2(t) = \frac{\sin \pi t - 2/\pi}{\sqrt{1/2 - 4/\pi^2}}$  with eigenvalues  $\lambda_1 = 0.25$ ,  $\lambda_2 = 0.0474$ , and  $\lambda_j = 0$  for  $j \geq 3$ . It is not hard to construct a kernel that satisfies the conditions in Theorem 5.2. For example, one can consider the kernel  $\mathcal{K}(r - t) = \sum_{j=1}^{+\infty} \lambda_j \{\cos(2\pi jr) \cos(2\pi jt) + \sin(2\pi jr) \sin(2\pi jt)\} = \sum_{j=1}^{+\infty} \lambda_j \cos(2\pi j(r - t))$  with  $\sum_{j=1}^{+\infty} \lambda_j = 1$  and  $\lambda_j = O(1/j^{19+\epsilon})$  for some  $\epsilon > 0$ . Then the asymptotic expansion (15) holds for the Wald statistic based on the difference kernel  $\mathcal{G}(r, t) = \mathcal{K}(r - t)$ .

Define the Parzen characteristic exponent

$$q = \max \left\{ q_0 : q_0 \in \mathbb{Z}^+, g_{q_0} = \lim_{x \rightarrow 0} \frac{1 - \mathcal{K}(x)}{|x|^{q_0}} < \infty \right\}.$$

For the Bartlett kernel  $q$  is 1; For the Parzen and QS kernels,  $q$  is equal to 2. Let  $c_1 = \int_{-\infty}^{+\infty} \mathcal{K}(x) dx$  and  $c_2 = \int_{-\infty}^{+\infty} \mathcal{K}^2(x) dx$ . We summarize the first and second order approximations for the distribution of studentized sample mean in the Gaussian location model based on both fixed- $b$  and small- $b$  asymptotics in Table 1 below. The formulae for the second order approximation under the small- $b$  asymptotics is from Velasco and Robinson (2001).

Table 1: Asymptotic comparison between the first and second order approximations based on fixed- $b$  and small- $b$  asymptotics.

Asymptotics	First order	Second order
Fixed- $b$	$P\left(\frac{B_1^2(1)}{Q(b)} \leq x\right)$	$P\left(\frac{B_1^2(1)}{Q(b)} \leq x\right) + \psi_{T,b}(x)$
Small- $b$	$G_1(x)$	$G_1(x) + (c_2 G_1''(x)x^2 - c_1 G_1'(x)x)b - \frac{g_q \sum_{h=-\infty}^{+\infty}  h ^q \gamma_X(h)}{\sigma^2(bT)^q} G_1'(x)x$

Note:  $Q(b) = \sum_{j=1}^{+\infty} \lambda_{j,b} v_j^2$ , where  $\{\lambda_{j,b}\}$  are the eigenvalues of the kernel  $\mathcal{K}((r - t)/b)$ .

**REMARK 5.6.** A few remarks are in order regarding Table 1. First of all, it is worth noting that  $P\left(\frac{B_1^2(1)}{Q(b)} \leq x\right) = G_1(x) + (c_2 G_1''(x)x^2 - c_1 G_1'(x)x)b + O(b^2)$  as  $b \rightarrow 0$  in Sun et al. (2008), which suggests that the fixed- $b$  limiting distribution captures the first two terms in the higher order asymptotic expansion under the small- $b$  asymptotics and thus provides a better approximation than the  $\chi_1^2$  approximation. Secondly, it is interesting to compare the second order asymptotic expansions under the fixed- $b$  asymptotics and small- $b$  asymptotics. We show in Proposition 5.3 that the high order expansion under fixed- $b$  asymptotics is consistent with the corresponding high order expansion under small- $b$  asymptotics as  $b$  approaches zero.

Because our fixed- $b$  expansion is established under the assumption that the eigenfunctions have mean zero, we shall consider the Wald statistic  $F_T(\infty)$  based on the demeaned kernel

$\tilde{\mathcal{G}}_b(r, t) = \mathcal{K}_b(r - t) - \int_0^1 \mathcal{K}_b(s - t) ds - \int_0^1 \mathcal{K}_b(r - p) dp + \int_0^1 \int_0^1 \mathcal{K}_b(s - p) ds dp$  for  $b \in (0, 1]$ . Let  $\{\tilde{\phi}_{j,b}\}$  and  $\{\tilde{\lambda}_{j,b}\}$  be the corresponding eigenfunctions and eigenvalues of  $\tilde{\mathcal{G}}_b(\cdot, \cdot)$ .

**PROPOSITION 5.3.** *Suppose  $\mathcal{K}(\cdot) : \mathbb{R} \rightarrow [0, 1]$  is symmetric, semi-positive definite, piecewise smooth with  $\mathcal{K}(0) = 1$  and  $\int_0^{+\infty} x\mathcal{K}(x)dx < \infty$ . The Parzen characteristic exponent of  $\mathcal{K}$  is no less than one. Further assume that*

$$\sup_{k \in \mathbb{N}} \left| \sum_{i=1}^k \tilde{\lambda}_{i,b} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) \right| = O \left( \sum_{i=1}^{+\infty} \tilde{\lambda}_{i,b} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) \right), \quad \text{as } b + 1/(bT) \rightarrow 0, \quad (16)$$

where  $\tilde{\xi}_{i,b}$  is defined by replacing  $\phi_j$  with  $\tilde{\phi}_{j,b}$  in the definition of  $\xi_i$ . Then under the assumption that  $\sigma^2 > 0$  and  $\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty$ , we have

$$\psi_{T,b}(x) = - \frac{g_q \sum_{h=-\infty}^{+\infty} |h|^q \gamma_X(h)}{\sigma^2 (bT)^q} G'_1(x) x (1 + o(1)) + O(1/T),$$

for fixed  $x \in \mathbb{R}$ , as  $b \rightarrow 0$  and  $bT \rightarrow +\infty$ .

In proposition 5.3, the condition (16) is not primitive and it requires that the bias for the LRV estimators based on the kernel  $\tilde{\mathcal{G}}_{k,b}(r, t) = \sum_{i=1}^k \tilde{\lambda}_{j,b} \tilde{\phi}_{j,b}(r) \tilde{\phi}_{j,b}(t)$  is at the same or smaller order of the bias for the LRV estimator based on  $\tilde{\mathcal{G}}_b(r, t)$ . This condition simplifies our technical arguments and it can be verified through a case-by-case study. As shown in proposition 5.3, the fixed- $b$  expansion is consistent with the small- $b$  expansion as  $b$  approaches zero and it is expected to be more accurate in terms of approximating the finite sample distribution when  $b$  is relatively large. Overall speaking, the above result suggests that the fixed- $b$  expansion provides a good approximation to the finite sample distribution which holds for a broad range of  $b$ .

## 6 Gaussian Dependent Bootstrap

Given the higher order expansions presented in Section 5, it seems natural to investigate if bootstrap can help to improve the first order approximation. To present the idea, we again limit our attention to the univariate Gaussian location model. Consider a consistent estimate of the covariance matrix of  $\{X_t\}_{t=1}^T$  which takes the form  $\hat{\Xi}(\omega; l) \in \mathbb{R}^{T \times T}$  with the  $(i, j)$ th element given by  $\omega_l(i - j) \hat{\gamma}_X(|i - j|)$  for  $i, j = 1, 2, \dots, T$ , where  $\omega$  is a kernel function with  $\omega_l(x) = \omega(x/l)$  and  $\hat{\gamma}_X(h) = \frac{1}{T} \sum_{t=1}^{T-h} (X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T)$  for  $h = 0, 1, 2, \dots, T-1$ . Estimating the covariance matrix of a stationary time series has been investigated by a few researchers. See Wu and Pourahmadi (2009) for the use of a banded sample covariance matrix and McMurry and Politis (2011) for a tapered version of the sample covariance matrix. In what follows, we shall consider the Bartlett kernel, i.e.,  $\omega(x) = (1 - |x|)\mathbf{I}\{|x| < 1\}$ , which guarantees to yield a semi-positive definite estimates, i.e.,  $\hat{\Xi}(\omega; l) \geq 0$ .

We now introduce a simple bootstrap procedure which can be shown to be second order correct. Suppose  $X_1^*, \dots, X_T^*$  is the bootstrap sample generated from  $N(0, \hat{\Xi}(\mathcal{K}, l))$ . It is easy to see that  $X_i^*$ 's are stationary and Gaussian conditional on the data. This is why we name this

bootstrap method “Gaussian Dependent Bootstrap”. There is a large literature on bootstrap for time series; see Lahiri (2003) for a review. However, most of the existing bootstrap methods do not deliver a conditionally normally distributed bootstrap sample. Since our higher order results are obtained under the Gaussian assumption, we need to generate Gaussian bootstrap sample in order for our expansion results to be useful.

Denote  $T_K^*$  the bootstrapped subsampling  $t$ -statistic obtained by replacing  $(X_1 - \mu_0, X_2 - \mu_0, \dots, X_T - \mu_0)$  with  $(X_1^*, X_2^*, \dots, X_T^*)$ . Define the bootstrapped projection vectors  $\xi_0^* = \frac{1}{\sqrt{T}} \sum_{j=1}^T X_j^*$  and  $\xi_i^* = \frac{1}{\sqrt{T}} \sum_{j=1}^T \phi_i^0(j/T) X_j^*$  for  $i = 1, \dots, +\infty$ . Let  $P^*$  be the bootstrap probability measure conditional on the data. The following theorems state the second order accuracy of the Gaussian dependent bootstrap in the univariate Gaussian location model.

**THEOREM 6.1.** *For the Gaussian location model, under the same conditions in Theorem 5.1 and  $1/l + l^3/T \rightarrow 0$ , we have*

$$\sup_{x \in [0, +\infty)} |P(|T_K| \leq x) - P^*(|T_K^*| \leq x)| = o_p(1/T). \quad (17)$$

**REMARK 6.1.** When  $K$  grows slowly with the sample size, the higher order expansions depend on the second order properties only through the quantities  $\sum_{h=-\infty}^{+\infty} |h|^k \gamma_X(h)$  with  $k = 0, 1, 2$  for the subsampling  $t$ -statistic and the Wald statistic based on the series variance estimator [see Proposition 5.2 and Theorem 4 of Sun (2011a)]. It suggests that the Gaussian dependent bootstrap also preserves the second order accuracy under the increasing-domain asymptotics provided that  $\sum_{h=-\infty}^{+\infty} |h|^3 \gamma_X(h) < \infty$ . A rigorous proof is omitted due to space limitation.

**THEOREM 6.2.** *For the Gaussian location model, under the assumptions in Theorem 5.2 and that  $1/l + l^3/T \rightarrow 0$ , we have*

$$\sup_{x \in [0, +\infty)} |P(F_T(\infty) \leq x) - P^*(F_T^*(\infty) \leq x)| = o_p(1/T), \quad (18)$$

where  $F_T^*(\infty) = \frac{(\xi_0^*)^2}{\sum_{j=1}^{+\infty} \lambda_j (\xi_j^*)^2}$  with  $\{\lambda_j\}_{j=1}^{+\infty}$  given in (13). Note that  $F_T^*(\infty) = (\xi_0^*)^2 / \hat{D}_T^*$ , where  $\hat{D}_T^* = T^{-1} \sum_{i,j=1}^T \mathcal{G}(i/T, j/T) (X_i^* - \bar{X}_T^*)(X_j^* - \bar{X}_T^*)$  and  $\bar{X}_T^*$  is the bootstrap sample mean.

The bootstrap-based autocorrelation robust testing procedures have been well studied in both econometric and statistical literatures under the increasing-smoothing asymptotics. In the statistical literature, Lahiri (1996) showed that for the studentized  $M$ -estimator, the ERP of the moving block bootstrap (MBB)-based testing procedure is of order  $o_p(T^{-1/2})$  which provides an asymptotic refinement to the normal approximation. Under the framework of the smooth function model, Götze and Künsch (1996) showed that the ERP for the MBB-based test is of order  $O_p(T^{-3/4+\epsilon})$  for any  $\epsilon > 0$  when the HAC estimator is constructed using the truncated kernel. Note that in the latter paper, the HAC estimator used in the studentized bootstrap statistic needs to take a different form from the original HAC estimator to achieve the higher order accuracy. Also see Lahiri (2007) for a recent contribution. In the econometric literature, the Edgeworth analysis for the block bootstrap has been conducted by Hall and Horowitz (1996),

Andrews (2002) and Inoue and Shintani (2006), among others, in the GMM framework. Within the increasing-smoothing asymptotic framework, it is still unknown whether the bootstrap can achieve an ERP of  $o_p(1/T)$  when a HAC covariance matrix estimator is used for studentization [see Härdle, Horowitz and Kreiss (2003)]. <sup>7</sup>

Within the fixed-smoothing asymptotic framework, Jansson (2004) established that the error of the fixed- $b$  approximation is of order  $O(\log(T)/T)$  for the Gaussian location model and the case  $b = 1$ , which was further refined by Sun et al. (2008) by dropping the  $\log(T)$  term. In the non-Gaussian setting, Gonçalves and Vogelsang (2011) showed that the fixed- $b$  approximation has an ERP of order  $o(T^{-1/2+\epsilon})$  for any  $\epsilon > 0$  when all moments exist. The latter authors further showed that the moving block bootstrap (with iid bootstrap as a special case) is able to replicate the fixed- $b$  limiting distribution and thus provides more accurate approximation than the normal approximation. However, because the exact form of the leading error term was not obtained in their studies, their results seem not directly applicable to show the higher order accuracy of bootstrap under the fixed- $b$  asymptotics. Using the asymptotic expansion results developed in Section 5, we show that the Gaussian dependent bootstrap can achieve an ERP of order  $o_p(1/T)$  under the Gaussian assumption. This appears to be the first result that shows the higher order accuracy of bootstrap under the fixed-smoothing asymptotics. Our result also provides a positive answer to the open question mentioned in Härdle, Horowitz and Kreiss (2003) that whether the bootstrap can achieve an ERP of  $o_p(1/T)$  in the dependence case when a HAC covariance matrix estimator is used for studentization<sup>8</sup>.

In the following, we conduct a small simulation study to compare and contrast the finite sample performance of the small- $b$  approximation, fixed- $b$  approximation, MBB, Gaussian dependent bootstrap (GDB), and the Edgeworth approximation derived by Velasco and Robinson (2001). Following the setup in Gonçalves and Vogelsang (2011), we consider the AR(1) model,

$$y_t = \rho y_{t-1} + \sqrt{1 - \rho^2} \varepsilon_t, \quad t = 1, 2, \dots, T, \quad (19)$$

with  $\{\varepsilon_t\}$  being a sequence of iid  $N(0, 1)$  or  $t(3)$  random variables. Consider the Wald statistic based on the HAC estimator with the Bartlett kernel and QS kernel for testing the null hypothesis  $E[y_t] = 0$  versus the alternative that  $E[y_t] \neq 0$  at 5% nominal level. Throughout the simulation we set  $T = 50$  and the number of Monte Carlo replications to be 1000. The bootstrap tests are based on 1000 replications for each sample. We implement the Edgeworth approximation in two ways (feasible and infeasible) as described in Gonçalves and Vogelsang (2011). The simulation results for  $b = 0.04, 0.06, 0.08, 0.1, 0.2, \dots, 1$  and  $\rho = -0.7, 0, 0.5, 0.9$  are summarized in Figures 2-3. We present the results for GDB with  $l = 5, 10$  and MBB with block size equal to 5. It is seen from the figures that the GDB is more accurate than the small- $b$  asymptotic approximation in most cases and improvement is often substantial especially for large  $b$ . In the

---

<sup>7</sup>Note that Hall and Horowitz (1996) and Andrews (2002) obtained the  $o_p(1/T)$  results but they assumed the uncorrelatedness of the moment conditions after finite lags.

<sup>8</sup>It is worth noting that our result is established under the fixed-smoothing asymptotics. It seems that in general the ERP of order  $o_p(1/T)$  cannot be achieved under the increasing-domain asymptotics.

dependent cases (e.g.,  $\rho = -0.7, 0.5$  and  $0.9$ ), the GDB tends to provide a refinement over the fixed- $b$  approximation for a proper bandwidth which is consistent with our theoretical findings. The improvement is apparent when the dependence is strong and  $b$  is small. In addition it is interesting to note that the GDB not only provides an improvement when the innovations are Gaussian but also in the case of  $t(3)$  distributed fat tailed innovations. The performance of GDB and MBB is in general quite close to each other. GDB tends to outperform MBB in the case of negative dependence whereas MBB delivers slightly better size in most cases when the dependence is positive. Finally, note that the performance of the feasible and infeasible Edgeworth approximation is similar to what has been described in Gonçalves and Vogelsang (2011) for the one-sided  $t$  test. Overall, the simulation results are consistent with those in Gonçalves and Vogelsang (2011), and they demonstrate the effectiveness of the proposed Gaussian dependent bootstrap in both Gaussian and non-Gaussian settings. The moving block bootstrap is expected to be second order accurate, as seen from its empirical performance, but a rigorous theoretical justification seems very difficult.

## 7 Conclusion

In this paper, we propose a general class of estimators to estimate the asymptotic covariance matrix of the GMM estimator in stationary time series models. Our proposal unifies a few existing covariance matrix estimators and reveals the connection among some recently developed fixed-smoothing approaches. First order asymptotic distribution of the Wald statistics with the general LRV estimator is obtained under the fixed-smoothing asymptotics. Under the framework of the Gaussian location model, we derive the Edgeworth expansion of the subsampling based  $t$ -statistic and the Wald statistic with the HAC estimator. Our work differs from the existing ones in two important aspects: (i) the expansion is derived under the fixed-smoothing asymptotics and the ERP of order  $O(1/T)$  is shown for a broad class of fixed-smoothing inference procedures; (ii) We obtain an explicit form for the leading error term, which is unavailable in the literature. An in-depth analysis of the behavior of the leading error term when the smoothing parameter grows with sample size (i.e.,  $K \rightarrow \infty$  in the subsampling  $t$ -statistic or  $b \rightarrow 0$  in the Wald statistic with the HAC estimator) shows the consistency of our results with the expansion results under the increasing-smoothing asymptotics. Building on these expansions, we further propose a new bootstrap method, the Gaussian dependent bootstrap, which provides a higher order correction than the first order fixed-smoothing approximation. Simulations results strongly suggest the relevance of our theory and the effectiveness of the Gaussian dependent bootstrap.

We mention a few directions that are worthy of future research. Firstly, it would be interesting to relax the Gaussian assumption in all the expansions we obtained in the paper. For non-Gaussian time series, Edgeworth expansions have been obtained by Götze and Kunsch (1996), Lahiri (2007, 2010), among others, for studentized statistics of a smooth function model under weak dependence assumption, but their results were derived under the increasing-smoothing asymptotics. For the location model and studentized sample mean, we conjecture that under

the fixed-smoothing asymptotics, the leading error term in the expansion of its distribution function involves the third and fourth order cumulants, which reflects the non-Gaussianness, and the order of the leading error term is  $O(T^{-1/2})$  instead of  $O(T^{-1})$ . Secondly, we expect that our expansion results will be useful in the optimal choice of the smoothing parameter, the kernel and its corresponding eigenvalues and eigenfunctions, for a given loss function. The optimal choice of the smoothing parameter has been addressed in Sun et al. (2008) using the expansion derived under the increasing-smoothing asymptotics. As the finite sample distribution is better approximated by the corresponding fixed-smoothing based approximations at either first or second order than its increasing-smoothing counterparts, the fixed-smoothing asymptotic theory proves to be more relevant in terms of explaining the finite sample results [see Gonçalves and Vogelsang (2011)]. Therefore, it might be worth reconsidering the choice of the optimal smoothing parameter under the fixed-smoothing asymptotics. Thirdly, we restrict our attention to the Gaussian location model when deriving the higher order expansions. It would be interesting to extend the results to the general GMM setting. A recent attempt by Sun (2010) for the HAC based inference seems to suggest this is feasible. Finally, the second order correctness of the moving block bootstrap for studentized sample mean, although suggested by the simulation results, is still an open but challenging topic for future research.

## References

- ANDERSON, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.
- ANDREWS, D. W. K. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817-858.
- (2002): “Higher-Order Improvements of a Computationally Attractive K-Step Bootstrap for Extreme Estimators,” *Econometrica*, 70, 119-162.
- BESTER C. A., T. G. CONLEY, AND C. B. HANSEN (2009): “Inference with Dependent Data Using Cluster Covariance Estimators,” Working paper, Department of Economics, University of Chicago.
- BILLINGSLEY, P. (1999): *Convergence of Probability Measures (Second Ed.)*. New York: Wiley.
- CHEN, Y., AND Z. QU (2010): “M Tests with a New Normalization Matrix,” Working paper, Department of Economics, Boston University.
- GONÇALVES, S., AND T. J. VOGELSANG (2011): “Block Bootstrap HAC Robust Tests: the Sophistication of the Naive Bootstrap,” *Econometric Theory*, 27, 745-791.
- GÖTZE, F., AND H. R. KÜNSCH (1996): “Second-Order Correctness of the Blockwise Bootstrap for Stationary Observations,” *Annals of Statistics*, 24, 1914-1933.



- GRENANDER, U., AND M. ROSENBLATT (1957): *Statistical Analysis of Stationary Time Series*. New York: Wiley.
- HANSEN, L. P. (1982): "Large Sample Properties of Generalized Methods of Moments Estimators," *Econometrica*, 50, 1029-1054.
- HALL, P., AND J. HOROWITZ (1996): "Bootstrap Critical Values for Tests Based on Generalized Method of Moment Estimators," *Econometrica*, 64, 891-916.
- HÄRDLE, W., J. HOROWITZ, AND J-P. KREISS (2003): "Bootstrap Methods for Time Series," *International Statistical Review*, 71, 435-459.
- HORN, R. A., AND C. R. JOHNSON (1986): *Matrix Analysis*. Cambridge University Press, Cambridge.
- IBRAGIMOV, R., AND U. K. MÜLLER (2010): " $t$ -statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business and Economic Statistics*, 28, 453-468.
- INOUE, A., AND M. SHINTANI (2006): "Bootstrapping GMM Estimators for Time Series," *Journal of Econometrics*, 133, 531-555.
- JANSSON, M. (2004): "On the Error of Rejection Probability in Simple Autocorrelation Robust Tests," *Econometrica*, 72, 937-946.
- KIEFER, N. M., AND T. J. VOGELSANG (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 1130-1164.
- KNESSL, C., AND J. B. KELLER (1991): "Asymptotic Properties of Eigenvalues of Integral Equations," *SIAM Journal on Applied Mathematics*, 21, 1130-1164.
- KUAN, C. M., AND W. M. LEE (2006): "Robust M Tests Without Consistent Estimation of the Asymptotic Covariance Matrix," *Journal of the American Statistical Association*, 101, 1264-1275.
- LAHIRI, S. N. (1996): "On Edgeworth Expansion and Moving Block Bootstrap for Studentized M-estimator in Multiple Linear Regression Models," *Journal of Multivariate Analysis*, 56, 42-59.
- (2003): *Resampling Methods for Dependent Data*. Springer, New York.
- (2007): "Asymptotic Expansions for Sums of Block-Variables Under Weak Dependence," *Annals of Statistics*, 35, 1324-1350.
- (2010): "Edgeworth Expansions for Studentized Statistics Under Weak Dependence," *Annals of Statistics*, 38, 388-434.
- LITTLE, S., AND J. B. READE (1984): "Eigenvalues of Analytic Kernels," *SIAM Journal on Mathematical Analysis*, 15, 133-136.



- LOBATO, I. N. (2001): “Testing That a Dependent Process is Uncorrelated,” *Journal of the American Statistical Association*, 96, 1066-1076.
- MCCULLOCH, C. E. (1982): “Symmetric Matrix Derivatives With Applications,” *Journal of the American Statistical Association*, 77, 679-682.
- MCMURRY, T. L., AND D. N. POLITIS (2010): “Banded and Tapered Estimates for Autocovariance Matrices and the Linear Process Bootstrap,” *Journal of Time Series Analysis*, 31, 471-482.
- NEWKEY, W. K., AND K. D. WEST (1987): “A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703-708.
- PHILLIPS, P. C. B. (2005): “HAC Estimation by Automated Regression,” *Econometric Theory*, 21, 116-142.
- SHAO, X. (2010): “A Self-normalized Approach to Confidence Interval Construction in Time Series,” *Journal of the Royal Statistical Society, Series, B.*, 72, 343-366.
- SUN, Y. (2010): “Let’s Fix It: Fixed- $b$  Asymptotics Versus Small- $b$  Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference,” Working paper, Department of Economics, UCSD.
- (2011a): “Autocorrelation Robust Inference Using Nonparametric Series Methods and Testing-Optimal Smoothing Parameter,” Working paper, Department of Economics, UCSD.
- (2011b): “Robust Trend Inference with Series Variance Estimator and Testing-Optimal Smoothing Parameter,” *Journal of Econometrics*, 164, 345-366.
- SUN, Y., P. C. B. PHILLIPS, AND S. JIN (2008): “Optimal Bandwidth Selection in Heteroscedasticity-Autocorrelation Robust Testing,” *Econometrica*, 76, 175-194.
- TURKINGTON, D. A. (2002): *Matrix Calculus and Zero-one Matrices: Statistical and Econometric Applications*. Cambridge University Press, Cambridge.
- VELASCO, C., AND P. M. ROBINSON (2001): “Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean,” *Econometric Theory*, 17, 497-539.
- WU, W. B., AND M. POURAHMADI (2009): “Banding Sample Autocovariance Matrices of Stationary Processes,” *Statistica Sinica*, 19, 1755-1768.

## 8 Appendix

### 8.1 Proof of the main results in section 4

*Proof of Theorem 4.1.* Define  $S_t(\hat{\theta}_T) = \frac{1}{T} \sum_{i=1}^t \hat{u}_i$ . Using the arguments at p. 5 of Sun (2011b), we can show that

$$\sqrt{T}S_{\lfloor Tr \rfloor}(\hat{\theta}_T) = \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor Tr \rfloor} \hat{u}_i \Rightarrow \Lambda B_p(r) := {}^d \Lambda(W_p(r) - rW_p(1)),$$

where  $\Lambda$  is invertible such that  $\Lambda\Lambda' = R(\theta_0)(G_0'W_0G_0)^{-1}G_0'W_0\Omega W_0G_0(G_0'W_0G_0)^{-1}R(\theta_0)'$  and  $W_p(r)$  is a  $p$ -dimensional vector of independent Brownian motions. Using summation by parts, we get

$$V_s = \frac{1}{bT} \sum_{t=1}^{T-1} \frac{[\phi_s\{t/(bT)\}] - \phi_s\{(t+1)/(bT)\}}{1/bT} \sqrt{T}S_t(\hat{\theta}_T) + \sqrt{T}\phi_s(1/b)S_T(\hat{\theta}_T),$$

where the last term disappears by recalling the fact that  $G_T(\hat{\theta}_T)'W_Tg_T(\hat{\theta}_T) = 0$ . By the continuous mapping theorem, we have

$$\begin{pmatrix} V_1 \\ \vdots \\ V_K \\ \sqrt{T}r(\hat{\theta}_T) \end{pmatrix} \rightarrow^d \begin{pmatrix} -\frac{\Lambda}{b} \int_0^1 \phi_1'(r/b)B_p(r)dr \\ \vdots \\ -\frac{\Lambda}{b} \int_0^1 \phi_K'(r/b)B_p(r)dr \\ \Lambda W_p(1) \end{pmatrix} =^d \begin{pmatrix} \Lambda \int_0^1 \tilde{\phi}_1(r/b)dW_p(r) \\ \vdots \\ \Lambda \int_0^1 \tilde{\phi}_K(r/b)dW_p(r) \\ \Lambda W_p(1) \end{pmatrix}.$$

Here we are using the fact that

$$\begin{aligned} -\frac{\Lambda}{b} \int_0^1 \phi_s'(r/b)B_p(r)dr &= \Lambda \int_0^1 \phi_s(r/b)dB_p(r) = \Lambda \int_0^1 \{\phi_s(r/b) - \int_0^1 \phi_s(r/b)dr\}dW_p(r) \\ &= \Lambda \int_0^1 \tilde{\phi}_s(r/b)dW_p(r), \end{aligned}$$

for  $1 \leq s \leq K$ . It is not hard to see that

$$\text{Cov} \left( \int_0^1 \tilde{\phi}_s(r/b)dW_p(r), \int_0^1 dW_p(r) \right) = 0$$

and

$$\text{Cov} \left( \int_0^1 \tilde{\phi}_s(r/b)dW_p(r), \int_0^1 \tilde{\phi}_t(r/b)dW_p(r) \right) = R_{st}I_p,$$

for  $1 \leq s, t \leq K$ , which implies

$$V = (V_1', V_2', \dots, V_K', \sqrt{T}r(\hat{\theta}_T)')' \rightarrow^d N(0, \tilde{R} \otimes \Lambda\Lambda'), \quad \text{where } \tilde{R} = \begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix}.$$

We thus get  $V^* = (L \otimes I_p)V \rightarrow^d N(0, LRL' \otimes \Lambda\Lambda') =^d N(0, I_K \otimes \Lambda\Lambda')$ . In other words,  $V^*$  is free of the effect of the basis functions asymptotically. Recall that  $\hat{D}_T = \sum_{s=1}^K \lambda_s V_s^* V_s^{*'}'$ , it is not hard to see that

$$F_T(K) = (\Lambda^{-1}\sqrt{T}r(\hat{\theta}_T))' \{\Lambda^{-1}\hat{D}_T(\Lambda^{-1})'\}^{-1} (\Lambda^{-1}\sqrt{T}r(\hat{\theta}_T))/p \rightarrow^d U_p' D_p^{-1} U_p/p,$$

where  $D_p = \sum_{j=1}^K \lambda_j \eta_j \eta_j'$  and  $\{\eta_j\}_{j=1}^K$  and  $U_p$  are iid with distribution  $N(0, I_p)$ . When  $\lambda_j = 1/K, j = 1, 2, \dots, K$ , it is straightforward to see that  $F_T(K) \rightarrow^d \frac{K}{K-p+1} F_{p, K-p+1}$ .  $\diamond$

From the proof of Theorem 4.1, we see that the choice of the normalization matrix  $R$  is somehow artificial and the limiting distribution is pivotal, for which the critical value is readily available. By introducing the orthogonalization matrix  $R$ , the orthonormal and zero mean assumptions for the basis functions can be dropped, compare Sun (2011b).

*Proof of Theorem 4.2.* Notice that  $\sqrt{Tr}(\hat{\theta}_T) \rightarrow^d N(c, \Lambda\Lambda')$  under the local alternatives. The result follows from the arguments in Theorem 4.1 and Theorem 5.2.2 in Anderson (2003).  $\diamond$

## 8.2 Proof of the main results in section 5.1

Consider the  $K+1$  dimensional multivariate normal density function which takes the form  $f(y, \Sigma) = (2\pi)^{-\frac{K+1}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}y'\Sigma^{-1}y\right)$ . We assume the  $(i, j)$ th element and the  $(j, i)$ th element of  $\Sigma$  are functionally unrelated. The results can be extended to the case where symmetric matrix elements are considered functionally equal [see e.g., McCulloch (1982)]. In the following, we use  $\otimes$  to denote the Kronecker product in matrix algebra and use  $\text{vec}$  to denote the operator that transforms a matrix into a column vector by stacking the columns of the matrix one underneath the other. For a vector  $y \in \mathbb{R}^{l \times 1}$  whose elements are differential functions of a vector  $x \in \mathbb{R}^{k \times 1}$ , we define  $\frac{\partial y}{\partial x}$  to be a  $k \times l$  matrix with the  $(i, j)$ th element being  $\frac{\partial y_j}{\partial x_i}$ . The notation  $u \asymp v$  represents  $u = O(v)$  and  $v = O(u)$ .

LEMMA 8.1.

$$\frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \Sigma) = \frac{f(y, \Sigma)}{2} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\}.$$

*Proof.* Note that

$$\begin{aligned} \frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \Sigma) &= (2\pi)^{-\frac{K+1}{2}} \left\{ \exp\left(-\frac{1}{2}y'\Sigma^{-1}y\right) \frac{\partial |\Sigma|^{-\frac{1}{2}}}{\partial \text{vec}(\Sigma)} + |\Sigma|^{-\frac{1}{2}} \frac{\partial}{\partial \text{vec}(\Sigma)} \exp\left(-\frac{1}{2}y'\Sigma^{-1}y\right) \right\} \\ &= (2\pi)^{-\frac{K+1}{2}} \left\{ -\frac{1}{2} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}y'\Sigma^{-1}y\right) \text{vec}(\Sigma^{-1}) \right. \\ &\quad \left. + \frac{1}{2} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}y'\Sigma^{-1}y\right) (\Sigma^{-1}y) \otimes (\Sigma^{-1}y) \right\} \\ &= \frac{f(y, \Sigma)}{2} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\}, \end{aligned}$$

where we have used the formulas  $\frac{\partial a'X^{-1}b}{\partial \text{vec}(X)} = -X^{-1}b \otimes (X^{-1})'a$  and  $\frac{\partial |X|^m}{\partial \text{vec}(X)} = m|X|^{m-1} \frac{\partial |X|}{\partial \text{vec}(X)} = m|X|^{m-1} \text{vec}((X^{-1})')$  [see Theorem 4.3 and Theorem 4.19 in Turkington (2005)].  $\diamond$

LEMMA 8.2.

$$\begin{aligned} \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\Sigma)}(y, \Sigma) &= \frac{1}{4} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\}' f(y, \Sigma) \\ &\quad - \frac{1}{2} \{(\Sigma^{-1}yy'\Sigma^{-1}) \otimes \Sigma^{-1} + \Sigma^{-1} \otimes (\Sigma^{-1}yy'\Sigma^{-1}) - \Sigma^{-1} \otimes \Sigma^{-1}\} f(y, \Sigma). \end{aligned}$$

*Proof.* From Lemma 8.1, we have

$$\begin{aligned} \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\Sigma)}(y, \Sigma) &= \frac{\partial}{\partial \text{vec}(\Sigma)} \left( \frac{f(y, \Sigma)}{2} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\} \right) \\ &= \left( \frac{\partial}{\partial \text{vec}(\Sigma)} \frac{f(y, \Sigma)}{2} \right) \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\}' \\ &\quad + \frac{f(y, \Sigma)}{2} \frac{\partial}{\partial \text{vec}(\Sigma)} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\} = I_1 + I_2. \end{aligned}$$

Again from Lemma 8.1, it is not hard to see that

$$I_1 = \frac{f(y, \Sigma)}{4} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\}'.$$

In view of Lemma 4.3 in Turkington (2005), we have

$$\frac{\partial \text{vec}(\Sigma^{-1}yy'\Sigma^{-1})}{\partial \text{vec}(\Sigma)} = \frac{\partial \text{vec}(\Sigma^{-1}y)}{\partial \text{vec}(\Sigma)} (y'\Sigma^{-1} \otimes I_{K+1}) + \frac{\partial \text{vec}(y'\Sigma^{-1})}{\partial \text{vec}(\Sigma)} (I_{K+1} \otimes y'\Sigma^{-1}).$$

Also by Theorem 4.3 in Turkington (2005), we get

$$\frac{\partial \text{vec}(\Sigma^{-1}y)}{\partial \text{vec}(\Sigma)} = -\Sigma^{-1}y \otimes \Sigma^{-1}; \quad \frac{\partial \text{vec}(y'\Sigma^{-1})}{\partial \text{vec}(\Sigma)} = -\Sigma^{-1} \otimes \Sigma^{-1}y,$$

which implies that

$$\frac{\partial \text{vec}(\Sigma^{-1}yy'\Sigma^{-1})}{\partial \text{vec}(\Sigma)} = -(\Sigma^{-1}yy'\Sigma^{-1}) \otimes \Sigma^{-1} - \Sigma^{-1} \otimes (\Sigma^{-1}yy'\Sigma^{-1}).$$

Further by Theorem 4.2 in Turkington (2005), we obtain  $\frac{\partial \text{vec}(\Sigma^{-1})}{\partial \text{vec}(\Sigma)} = -\Sigma^{-1} \otimes \Sigma^{-1}$ . The conclusion thus follows directly from the above derivation.  $\diamond$

**LEMMA 8.3.** *Let  $\{\Sigma_T\} \subset \mathbb{R}^{(K+1) \times (K+1)}$  be a sequence of positive definite matrices with  $K+1 \leq T$ . If  $K$  is fixed with respect to  $T$  and  $\|\Sigma_T - \Sigma\|_2 = O(1/T)$  for a positive definite matrix  $\Sigma$ , then we have*

$$\|\Sigma_T^{-1} - \Sigma^{-1}\|_2 = O(1/T).$$

*Proof.* Let  $\Sigma_T = \Sigma + R_T$  with  $\|R_T\|_2 = O(1/T)$ . For sufficiently large  $T$ , we have  $\|\Sigma^{-1}R_T\|_2 \leq \|\Sigma^{-1}\|_2 \|R_T\|_2 < 1$ . By the last equation at p. 355 of Horn and Johnson (1986), we have

$$\|\Sigma_T^{-1} - \Sigma^{-1}\|_2 \leq \frac{\|\Sigma^{-1}\|_2^2 \|R_T\|_2}{1 - \|\Sigma^{-1}R_T\|_2} = O(1/T).$$

$\diamond$

**LEMMA 8.4.** *Let  $\tilde{\Sigma}_T(y)$  be a  $(K+1) \times (K+1)$  positive symmetric matrix which depends on  $y \in \mathbb{R}^{K+1}$ . Assume that  $\sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T(y) - \Sigma\|_2 \leq \|\Sigma_T - \Sigma\|_2 = O(1/T)$  for a positive definite matrix  $\Sigma$ . Let  $R_T = \Sigma_T - \Sigma$ . If  $K$  is fixed with respect to  $T$ , we have*

$$\int_{y \in \mathbb{R}^{K+1}} \left| \text{vec}(R_T)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\Sigma)}(y, \tilde{\Sigma}(y)) \text{vec}(R_T) \right| dy = O(1/T^2).$$

*Proof.* Let  $\tilde{R}_T(y) = \tilde{\Sigma}_T(y) - \Sigma$ . Note that  $\sup_{y \in \mathbb{R}^{K+1}} \|\Sigma^{-1}\tilde{R}_T(y)\|_2 \leq \|\Sigma^{-1}\|_2 \sup_{y \in \mathbb{R}^{K+1}} \|\tilde{R}_T(y)\|_2 \leq \|\Sigma^{-1}\|_2 \|\Sigma_T - \Sigma\|_2 < 1$ , for large enough  $T$ . By using the same arguments in Lemma 8.3, we have  $\sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T^{-1}(y) - \Sigma^{-1}\|_2 = O(1/T)$ . Therefore, when  $T$  is sufficiently large, we have  $y'(\tilde{\Sigma}_T^{-1}(y) - \Sigma^{-1}/2)y = y'(\tilde{\Sigma}_T^{-1}(y) - \Sigma^{-1})y + y'\Sigma^{-1}y/2 \geq (\lambda_{\min}(\Sigma^{-1})/2 - \|\tilde{\Sigma}_T^{-1}(y) - \Sigma^{-1}\|_2)\|y\|^2 \geq 0$  for all  $y$ , where  $\lambda_{\min}(\Sigma^{-1})$  denotes the smallest eigenvalue of  $\Sigma^{-1}$ . On the other hand, for sufficiently large  $T$ , we have  $\sup_{y \in \mathbb{R}^{K+1}} |\tilde{\Sigma}_T(y)|^{-1} = \sup_{y \in \mathbb{R}^{K+1}} |\tilde{\Sigma}_T^{-1}(y)| \leq \sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T^{-1}(y)\|_2^{K+1} \leq (\|\Sigma^{-1}\|_2 + \sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T^{-1}(y) - \Sigma^{-1}\|_2)^{K+1} \leq C\|\Sigma\|^{-1}$  with  $C > 0$ . Combining the above arguments, we get  $f(y, \tilde{\Sigma}_T(y)) \leq C\|\Sigma\|^{-1/2} \exp(-y'\Sigma^{-1}y/4) \leq Cf(y, 2\Sigma)$  for all  $y$ . When  $K$  is fixed,  $\|\cdot\|_2$  and  $\|\cdot\|_\infty$  are equivalent, which implies  $\sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T(y)^{-1} - \Sigma^{-1}\|_\infty = O(1/T)$ . Since the elements of  $\tilde{\Sigma}_T^{-1}(y)$  are

uniformly bounded for all  $y$ , in view of Lemma 8.2, it is straightforward to see

$$\left| \text{vec}(R_T)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \text{vec}(\Sigma)}(y, \tilde{\Sigma}_T(y)) \text{vec}(R_T) \right| \leq Cp(y)f(y, 2\Sigma)/T^2,$$

where  $p(y)$  is a polynomial of degree 4. The conclusion follows by noting that  $\int p(y)f(y, 2\Sigma)dy < \infty$ .  $\diamond$

*Proof of Theorem 5.1.* For the convenience of our presentation, we ignore the functional symmetry of the covariance matrix in the proof. With some proper modifications, we can extend the results to the case where the functional symmetry is taken into consideration. let  $|\mathcal{G}_1| = |\mathcal{G}_2| = \dots = |\mathcal{G}_K| = q$ . Define  $Y_i = \sqrt{q}(\hat{\mu}_i - \mu_0)$ , and  $\bar{Y} = \frac{1}{K} \sum_{i=1}^K Y_i$  and  $S_Y^2 = \frac{1}{K-1} \sum_{i=1}^K (Y_i - \bar{Y})^2$  as the sample mean and sample variance of  $\{Y_i\}_{i=1}^K$  respectively. Note that  $T_K(Y) = \sqrt{K}\bar{Y}/S_Y$ , where  $Y = (Y_1, Y_2, \dots, Y_K)'$ . Simple algebra yields that

$$\sigma_{ij} := \text{Cov}(Y_i, Y_j) = \sum_{h=1-q}^{q-1} \left( \frac{q-|h|}{q} \right) \gamma_X(h - (j-i)q).$$

Notice that  $Y$  follows a normal distribution with mean zero and covariance matrix  $\Sigma_T$ , where  $\Sigma_T = (\sigma_{ij})_{i,j=1}^K$ . The density function of  $Y$  is given by,

$$f(y, \Sigma_T) = (2\pi)^{-K/2} |\Sigma_T|^{-1/2} \exp \left( -\frac{1}{2} y' \Sigma_T^{-1} y \right).$$

Under the assumption  $\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty$ , it is straightforward to see that  $\|\Sigma_T - \sigma^2 I_K\|_2 = O(1/T)$ . Taking a Taylor expansion of  $f(y, \Sigma_T)$  around elements of the matrix  $\sigma^2 I_K$ , we have

$$\begin{aligned} f(y, \Sigma_T) &= f(y, \sigma^2 I_K) + \left\{ \frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \sigma^2 I_K) \right\}' \text{vec}(\Sigma_T - \sigma^2 I_K) \\ &\quad + \text{vec}(\Sigma_T - \sigma^2 I_K)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \text{vec}(\Sigma)}(y, \tilde{\Sigma}_T(y)) \text{vec}(\Sigma_T - \sigma^2 I_K), \end{aligned}$$

where  $\sup_{y \in \mathbb{R}^K} \|\tilde{\Sigma}_T(y) - \sigma^2 I_K\|_2 \leq \|\Sigma_T - \sigma^2 I_K\|_2 = O(1/T)$ . By Lemma 8.1 and Lemma 8.4, we get

$$\frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \sigma^2 I_K) = f(y, \sigma^2 I_K) \left\{ -\frac{1}{2\sigma^2} \text{vec}(I_K) + \frac{1}{2\sigma^4} y \otimes y \right\},$$

and

$$\int_{y \in \mathbb{R}^K} \left| \text{vec}(\Sigma_T - \sigma^2 I_K)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \text{vec}(\Sigma)}(y, \tilde{\Sigma}_T(y)) \text{vec}(\Sigma_T - \sigma^2 I_K) \right| dy = O\left(\frac{1}{T^2}\right), \quad (20)$$

which imply that

$$\begin{aligned} f(y, \Sigma_T) &= f(y, \sigma^2 I_K) \left\{ 1 - \frac{1}{2\sigma^2} \sum_{i=1}^K (\sigma_{ii} - \sigma^2) \right\} + \frac{1}{2\sigma^4} f(y, \sigma^2 I_K) \sum_{i=1}^K \sum_{j=1}^K (\sigma_{ij} - \sigma^2 \delta_{ij}) y_i y_j + R(y) \\ &= g(y, \sigma^2 I_K) + R(y), \end{aligned}$$

where  $g$  denotes the major term,  $R(y)$  is the remainder term and  $\delta_{ij} = \mathbf{I}\{i = j\}$  is the kronecker's delta. Define  $\tilde{\Psi}_K(x) = \int_{\{|T_K(y)| > x\}} g(y, \sigma^2 I_K) dy$ . By (20), we see that

$$\sup_{x \in \mathbb{R}} \left| \int_{\{|T_K(y)| > x\}} f(y, \Sigma_T) dy - \tilde{\Psi}_K(x) \right| \leq \int_{\mathbb{R}^K} |R(y)| dy = O(1/T^2).$$

It follows from some simple calculation that

$$\tilde{\Psi}_K(x) = \left\{ 1 - \frac{1}{2\sigma^2} \sum_{i=1}^K (\sigma_{ii} - \sigma^2) \right\} P(|t_{K-1}| > x) + \frac{1}{2\sigma^2} (J_1 + J_2),$$

where

$$J_1 = \sum_{i=1}^K (\sigma_{ii} - \sigma^2) E[\mathbf{I}\{|\tilde{T}_K(v)| > x\} v_i^2], \quad J_2 = \sum_{i \neq j} \sigma_{ij} E[\mathbf{I}\{|\tilde{T}_K(v)| > x\} v_i v_j].$$

Here  $\{v_i\}_{i=1}^K$  are iid standard normal random variables and  $\tilde{T}_K(v) = \sqrt{K}\bar{v}/S_v$  is the  $t$  statistic based on  $\{v_i\}$  with  $\bar{v} = \frac{1}{K} \sum_{i=1}^K v_i$  and  $S_v^2 = \frac{1}{K-1} \sum_{i=1}^K (v_i - \bar{v})^2$ . Let  $U = K\bar{v}^2$  and  $D = (K-1)S_v^2$ . Then  $U \sim \chi_1^2$ ,  $D \sim \chi_{K-1}^2$ , and  $U$  and  $D$  are independent. We define that

$$\begin{aligned} E[\mathbf{I}\{|\tilde{T}_K(v)| > x\} v_i^2] &= \frac{1}{K} E[\mathbf{I}\{|\tilde{T}_K(v)| > x\} \sum_{i=1}^K v_i^2] \\ &= \frac{1}{K} E[\mathbf{I}\{|\tilde{T}_K(v)| > x\} U] + \frac{1}{K} E[\mathbf{I}\{|\tilde{T}_K(v)| > x\} D] \\ &= \frac{1}{K} E \left[ U G_{K-1} \left( \frac{(K-1)U}{x^2} \right) \right] + \frac{1}{K} E \left[ D - D G_1 \left( \frac{Dx^2}{K-1} \right) \right], \end{aligned}$$

and

$$\begin{aligned} E[\mathbf{I}\{|\tilde{T}_K(v)| > x\} v_i v_j] &= \frac{1}{K(K-1)} E[\mathbf{I}\{|\tilde{T}_K(v)| > x\} \sum_{i \neq j} v_i v_j] \\ &= \frac{1}{K-1} E[\mathbf{I}\{|\tilde{T}_K(v)| > x\} U] - \frac{1}{K(K-1)} E[\mathbf{I}\{|\tilde{T}_K(v)| > x\} \sum_{i=1}^K v_i^2] \\ &= \frac{1}{K} E \left[ U G_{K-1} \left( \frac{(K-1)U}{x^2} \right) \right] - \frac{1}{K(K-1)} E \left[ D - D G_1 \left( \frac{Dx^2}{K-1} \right) \right]. \end{aligned}$$

We then have

$$\begin{aligned} P(|T_K| > x) &= \tilde{\Psi}_K(x) + O(1/T^2) = \{1 - \alpha\} P(|t_{K-1}| > x) + \beta E \left[ U G_{K-1} \left( \frac{(K-1)U}{x^2} \right) \right] \\ &\quad + \tau \left\{ K - 1 - E \left[ D G_1 \left( \frac{Dx^2}{K-1} \right) \right] \right\} + O(1/T^2), \end{aligned} \tag{21}$$

uniformly for  $x \in \mathbb{R}$ , where the coefficients are given by

$$\alpha = \frac{1}{2\sigma^2} \sum_{i=1}^K (\sigma_{ii} - \sigma^2) = -\frac{K^2 B}{2\sigma^2 T} + O(1/T^2), \quad \beta = \frac{1}{2K\sigma^2} \sum_{i=1}^K \sum_{j=1}^K (\sigma_{ij} - \delta_{ij}\sigma^2) = -\frac{B}{2\sigma^2 T} + O(1/T^2),$$

and

$$\tau = \frac{1}{2K\sigma^2} \sum_{i=1}^K (\sigma_{ii} - \sigma^2) - \frac{1}{2K(K-1)\sigma^2} \sum_{i \neq j} \sigma_{ij} = -\frac{(K+1)B}{2\sigma^2 T} + O(1/T^2).$$

The conclusion thus follows from equation (21).  $\diamond$

*Proof of Proposition 5.1.* Note first that

$$\Upsilon(x, K)/K = -K P(|t_{K-1}| \leq x) + \frac{K+1}{K} E \left[ \chi_{K-1}^2 G_1 \left( \frac{\chi_{K-1}^2}{K-1} x^2 \right) \right] + O(1/K).$$

Using the fact that  $P(|t_{K-1}| \leq x) = G_1(x^2) + \frac{1}{K-1}x^4G_1''(x^2) + O(1/K^2)$ , we get

$$\begin{aligned}\Upsilon(x, K)/K &= -KG_1(x^2) - \frac{K}{K-1}x^4G_1''(x^2) + \frac{K+1}{K}E\left[\chi_{K-1}^2\left\{G_1(x^2)\right.\right. \\ &\quad \left.\left.\left(\frac{\chi_{K-1}^2}{K-1} - 1\right)x^2G_1'(x^2) + \frac{1}{2}\left(\frac{\chi_{K-1}^2}{K-1} - 1\right)^2x^4G_1''(x^2)\right\}\right] + O(1/K) \\ &= 2x^2G_1'(x^2) + O(1/K).\end{aligned}$$

◇

*Proof of Proposition 5.2.* Recall that  $q = T/K$  is assumed to be an integer. Using the notation in the proof of Theorem 5.1, let  $S_Y^2 = \frac{1}{K-1} \sum_{i=1}^K (Y_i - \bar{Y})^2 = \frac{1}{K-1} \{\sum_{i=1}^K Y_i^2 - K(\bar{Y})^2\}$ . Notice that

$$\begin{aligned}\text{cov}(Y) &= \begin{pmatrix} \sigma^2 - B/q & B/(2q) & 0 & \dots & 0 \\ B/(2q) & \sigma^2 - B/q & B/(2q) & \dots & 0 \\ & & \dots & & \\ 0 & 0 & \dots & B/(2q) & \sigma^2 - B/q \end{pmatrix}_{K \times K} \\ &\quad + O(1/q^2)l_K l_K' \\ &= \sigma^2 I_K + \frac{B}{2q} \begin{pmatrix} -2 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & \dots & 0 \\ & & \dots & & \\ 0 & 0 & \dots & 1 & -2 \end{pmatrix}_{K \times K} + O(1/q^2)l_K l_K' \\ &= \sigma^2 I_K + \frac{B}{2q} M + O(1/q^2)l_K l_K',\end{aligned}$$

where  $l_K' = (1, 1, \dots, 1)_{1 \times K}$  and the summation of all the  $O(1/q^2)$  is of order  $O(K/q^2)$ . Because

$$E[Y_i^2] = \sum_{h=1-q}^{q-1} \left( \frac{q-|h|}{q} \right) \gamma_X(h) = \sigma^2 - B/q + O(1/q^2),$$

and

$$E[\bar{Y}^2] = \frac{1}{K^2} \sum_{i,j=1}^K E[Y_i Y_j] = \frac{1}{K^2} \{K\sigma^2 - B/q + O(K/q^2)\} = \sigma^2/K + O(1/(K^2q)) + O(1/(Kq^2)),$$

we obtain

$$E[S_Y^2] - \sigma^2 = \frac{K}{K-1} \{\sigma^2 - B/q - \sigma^2/K + o(1/T)\} - \sigma^2 = -B/q + O(1/T).$$

Consider the covariance matrix of  $\tilde{Y}' = (Y_1 - \bar{Y}, Y_2 - \bar{Y}, \dots, Y_K - \bar{Y})$ . It is easy to see that  $\tilde{Y} = (I_K - l_K l_K'/K)Y = H_K Y$ , where  $H_K = I_K - l_K l_K'/K$  is an idempotent matrix. Ignoring the  $O(1/q^2)$  order term in  $\text{cov}(Y)$ , we have

$$\begin{aligned}(c_{ij})_{i,j=1}^K &:= \text{cov}(\tilde{Y}) = H_K \text{cov}(Y) H_K \approx H_K \{\sigma^2 I_K + BM/(2q)\} H_K \\ &= \sigma^2 H_K + \frac{B}{2q} H_K M H_K = \sigma^2 H_K + \frac{B}{2q} \left( M - \frac{1}{K} A - \frac{2}{K^2} l_K l_K' \right),\end{aligned}$$



where

$$A = \begin{pmatrix} -2 & -1 & -1 & \dots & -2 \\ -1 & 0 & 0 & \dots & -1 \\ & & \dots & & \\ -1 & 0 & 0 & \dots & -1 \\ -2 & -1 & -1 & \dots & -2 \end{pmatrix}_{K \times K}.$$

Since  $\tilde{Y}$  is Gaussian, we get

$$E[S_Y^4] = \frac{1}{(K-1)^2} \sum_{i,j=1}^K E[(Y_i - \bar{Y})^2 (Y_j - \bar{Y})^2] = \frac{1}{(K-1)^2} \sum_{i,j=1}^K (c_{ii}c_{jj} + 2c_{ij}^2),$$

where  $c_{ii} = (1 - \frac{1}{K})\sigma^2 - \frac{B}{q} + O(1/T)$  and  $c_{ij} = -\frac{1}{K}\sigma^2 + \frac{B}{2q}\mathbf{I}\{|i-j|=1\} + O(1/T)$ , for  $i \neq j$ . It implies that

$$\begin{aligned} \sum_{i,j=1}^K c_{ij}^2 &= \sum_{i=1}^K c_{ii}^2 + \sum_{|i-j|=1} c_{ij}^2 + \sum_{|i-j|>1} c_{ij}^2 = K \left(1 - \frac{1}{K}\right)^2 \sigma^4 + \frac{KB^2}{q^2} - \frac{2(K-1)B}{q} \sigma^2 \\ &\quad + 2(K-1) \left( \frac{\sigma^4}{K^2} + \frac{B^2}{4q^2} - \frac{\sigma^2 B}{Kq} \right) + \frac{(K-1)(K-2)}{K^2} \sigma^4 + O(1/q) \\ &= (K-1)\sigma^4 + O(K/q), \end{aligned}$$

and

$$\sum_{i,j=1}^K c_{ii}c_{jj} = K^2 c_{11}^2 + O(K/q) = (K-1)^2 \sigma^4 - \frac{2BK(K-1)\sigma^2}{q} + O(K/q).$$

Therefore we get

$$E[S_Y^4] = \frac{K+1}{K-1} \sigma^4 - \frac{2BK\sigma^2}{(K-1)q} + O(1/T),$$

which implies

$$\text{var}(S_Y^2) = \frac{K+1}{K-1} \sigma^4 - \frac{2BK\sigma^2}{(K-1)q} - (\sigma^2 - B/q)^2 + O(1/T) = \frac{2\sigma^4}{K-1} + O(1/T).$$

Let  $\mathbf{X} = (X_1, X_2, \dots, X_T)'$ ,  $\hat{\mu}_{GLS} = (l_T' \text{cov}(\mathbf{X})^{-1} l_T)^{-1} l_T' \text{cov}(\mathbf{X})^{-1} \mathbf{X}$  and  $\sigma_{GLS}^2 = T \text{var}(\hat{\mu}_{GLS}) = T(l_T' \text{cov}(\mathbf{X})^{-1} l_T)^{-1}$ . Note that  $\hat{\mu}_{GLS} - \mu_0$  is independent of  $S_Y$  and  $\sigma_{GLS}^2 = \sigma^2 + O(1/T)$  [see Grenander and Rosenblatt (1957)]. Using similar arguments in Lemma 1 of Sun (2011b), we have

$$\begin{aligned} P(|T_K| \leq x) &= P\left(\frac{T(\hat{\mu}_{GLS} - \mu_0)^2 / \sigma_{GLS}^2}{S_Y^2 / \sigma_{GLS}^2} \leq x^2\right) + O(1/T) \\ &= E[G_1(S_Y^2 x^2 / \sigma^2)] + O(1/T) \\ &= G_1(x^2) + \frac{x^2}{\sigma^2} G_1'(x^2) E[S_Y^2 - \sigma^2] + \frac{x^4 G_1''(x^2)}{2\sigma^4} E[(S_Y^2 - \sigma^2)^2] + O(1/T) \\ &= G_1(x^2) - \frac{BK}{T\sigma^2} x^2 G_1'(x^2) + \frac{1}{K-1} x^4 G_1''(x^2) + O(1/T). \end{aligned}$$

◇

### 8.3 Proof of the main results in section 5.2

We first establish a high order expansion for Wald statistic based on the general LRV estimators considered in section 3. Let  $\xi = (\xi_0, \xi_1, \dots, \xi_K)$  with  $\xi_0 = \frac{1}{\sqrt{T}} \sum_{i=1}^T (X_i - \mu_0)$  and  $\xi_j = \frac{1}{\sqrt{T}} \sum_{i=1}^T \phi_j^0(i/T) X_i$  for  $j = 1, 2, \dots, K$ , and  $\Sigma_\xi$  be the covariance matrix of  $\xi$ . We first show that the convergence rate of  $\Sigma_\xi$  is of order  $O(1/T)$  when the number of basis functions  $K$  is fixed.

**LEMMA 8.5.** *Assume the basis functions  $\{\phi_s(t)\}_{s=1}^K$  are bounded with finite discontinuous points and satisfy  $\sup_{\alpha \in (0,1]} \left\{ \left| \frac{1}{\alpha} \int_0^{1-\alpha} \tilde{\phi}_s(x) \{\tilde{\phi}_r(x+\alpha) - \tilde{\phi}_r(x)\} dx \right| + \left| \frac{1}{\alpha} \int_\alpha^1 \tilde{\phi}_s(x) \{\tilde{\phi}_r(x-\alpha) - \tilde{\phi}_r(x)\} dx \right| \right\} < \infty$ , for  $1 \leq s, r \leq K$ . Recall that  $\tilde{\phi}_s(t) = \phi_s(t) - \int_0^1 \phi_s(t) dt$ . If  $\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty$  and  $K$  is fixed, then we have  $\|\Sigma_\xi - \sigma^2 \tilde{R}\|_\infty = O(1/T)$ .*

*Proof of Lemma 8.5.* For  $s = 1, 2, \dots, K$ , we have

$$\text{cov}(\xi_0, \xi_s) = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \gamma_X(j-i) \phi_s^0\left(\frac{j}{T}\right) = \frac{1}{T} \sum_{h=1-T}^{T-1} \gamma_X(h) \sum_{1 \leq i, h+i \leq T} \phi_s^0\left(\frac{h+i}{T}\right).$$

Simple algebra gives us

$$\frac{1}{T} \sum_{1 \leq i, h+i \leq T} \phi_s^0\left(\frac{h+i}{T}\right) = \begin{cases} \frac{h}{T^2} \sum_{i=1}^T \phi_s(i/T) - \frac{1}{T} \sum_{i=1}^h \phi_s(i/T), & h > 0; \\ \frac{|h|}{T^2} \sum_{i=1}^T \phi_s(i/T) - \frac{1}{T} \sum_{i=T-|h|+1}^T \phi_s(i/T), & h < 0. \end{cases}$$

It implies that

$$\text{cov}(\xi_0, \xi_s) = \frac{1}{T} \int_0^1 \phi_s(t) dt \sum_{h=-\infty}^{+\infty} |h| \gamma_X(h) - \frac{1}{T} \sum_{0 < h < T} \gamma_X(h) \left\{ \sum_{i=1}^h \phi_s(i/T) + \sum_{i=T-h+1}^T \phi_s(i/T) \right\} + O(1/T^2). \quad (22)$$

Note that the second term on the right hand side of (22) is of order  $O(1/T)$  because the basis functions  $\{\phi_s(t)\}$  are bounded. Consider the covariance between  $\xi_s$  and  $\xi_r$  with  $1 \leq s, r \leq K$ . Straightforward calculation yields

$$\begin{aligned} \text{cov}(\xi_s, \xi_r) &= \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \phi_s^0\left(\frac{i}{T}\right) \phi_r^0\left(\frac{j}{T}\right) \gamma_X(i-j) \\ &= \frac{1}{T} \sum_{h=1}^{T-1} \sum_{1 \leq j, j+h \leq T} \phi_s^0\left(\frac{j+h}{T}\right) \phi_r^0\left(\frac{j}{T}\right) \gamma_X(h) + \frac{1}{T} \sum_{h=1-T}^{-1} \sum_{1 \leq j, j+h \leq T} \phi_s^0\left(\frac{j+h}{T}\right) \phi_r^0\left(\frac{j}{T}\right) \gamma_X(h) \\ &\quad + \gamma_X(0) \frac{1}{T} \sum_{j=1}^T \phi_s^0\left(\frac{j}{T}\right) \phi_r^0\left(\frac{j}{T}\right) = I_1 + I_2 + I_3. \end{aligned}$$

Notice that

$$\frac{1}{T} \sum_{1 \leq j \leq T} \phi_s^0\left(\frac{j}{T}\right) \phi_r^0\left(\frac{j}{T}\right) = \int_0^1 \tilde{\phi}_s(t) \tilde{\phi}_r(t) dt + C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t)) = R_{sr} + C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t)), \quad (23)$$

where  $C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t))$  is of order  $O(1/T)$ . It is not hard to see that

$$I_1 = \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left[ \sum_{j=1}^{T-h} \phi_r^0\left(\frac{j}{T}\right) \left\{ \phi_s^0\left(\frac{j+h}{T}\right) - \phi_s^0\left(\frac{j}{T}\right) \right\} \right], \quad \text{say } J_{1,T}$$

$$+ \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left\{ \sum_{j=1}^T \phi_r^0\left(\frac{j}{T}\right) \phi_s^0\left(\frac{j}{T}\right) - \sum_{j=T-h+1}^T \phi_r^0\left(\frac{j}{T}\right) \phi_s^0\left(\frac{j}{T}\right) \right\},$$

and

$$I_2 = \frac{1}{T} \sum_{h=1-T}^{-1} \gamma_X(h) \left[ \sum_{j=1+|h|}^T \phi_r^0\left(\frac{j}{T}\right) \left\{ \phi_s^0\left(\frac{j+h}{T}\right) - \phi_s^0\left(\frac{j}{T}\right) \right\} \right], \quad \text{say } J_{2,T}$$

$$+ \frac{1}{T} \sum_{h=1-T}^{-1} \gamma_X(h) \left\{ \sum_{j=1}^T \phi_r^0\left(\frac{j}{T}\right) \phi_s^0\left(\frac{j}{T}\right) - \sum_{j=1}^{|h|} \phi_r^0\left(\frac{j}{T}\right) \phi_s^0\left(\frac{j}{T}\right) \right\}.$$

Using (23), we have

$$\begin{aligned} \text{cov}(\xi_s, \xi_r) &= \left\{ \sigma^2 - \sum_{|h| \geq T} \gamma_X(h) \right\} \left\{ R_{sr} + C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t)) \right\} - \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left\{ \sum_{j=1}^h \phi_r^0\left(\frac{j}{T}\right) \phi_s^0\left(\frac{j}{T}\right) \right. \\ &\quad \left. + \sum_{j=T-h+1}^T \phi_r^0\left(\frac{j}{T}\right) \phi_s^0\left(\frac{j}{T}\right) \right\} + J_{1,T} + J_{2,T}. \end{aligned} \quad (24)$$

Under the assumption that  $\sup_{\alpha \in [0,1]} \left| \frac{1}{\alpha} \int_0^{1-\alpha} \tilde{\phi}_r(x)(\tilde{\phi}_s(x+\alpha) - \tilde{\phi}_s(x)) dx \right| < \infty$ , it is straightforward to see that

$$\begin{aligned} |J_{1,T}| &\leq \frac{1}{T} \sum_{h=1}^{T-1} |h \gamma_X(h)| \sup_{1 \leq h \leq T} \left| \frac{1}{h} \sum_{j=1}^{T-h} \phi_r^0\left(\frac{j}{T}\right) \left\{ \phi_s^0\left(\frac{j+h}{T}\right) - \phi_s^0\left(\frac{j}{T}\right) \right\} \right| \\ &\leq \frac{C}{T} \sum_{h=1}^{T-1} |h \gamma_X(h)| \left\{ \sup_{\alpha \in [0,1]} \left| \frac{1}{\alpha} \int_0^{1-\alpha} \tilde{\phi}_r(x)(\tilde{\phi}_s(x+\alpha) - \tilde{\phi}_s(x)) dx \right| \right\}, \end{aligned}$$

which implies that  $J_{1,T} = O(1/T)$ . The same argument applies to  $J_{2,T}$ . The proof is then complete.  $\diamond$

The assumption regarding the basis functions in Lemma 8.5 is mild. If  $\{\phi_j(t)\}_{j=1}^K$  are lipschitz continuous of order one, then the assumption is satisfied. It is easy to check that the condition holds for all three types of basis functions considered in Section 3. Recall the definition of  $R$  and  $L$  from Section 3, and let  $\tilde{R} = \text{diag}(1, R) = (\tilde{R}_{ij})_{i,j=0}^K$  and  $\tilde{L} = \tilde{R}^{1/2} = (\tilde{L}_{ij})_{i,j=0}^K$ . Define

$$\Phi_{T,K}(x) = Q_{1,K}(x) + \frac{1}{2\sigma^2} \sum_{i,j=0}^K \sum_{s=0}^K \tilde{L}_{si} \tilde{L}_{sj} (\text{cov}(\xi_i, \xi_j) - \sigma^2 \tilde{R}_{ij}) E[\mathbf{I}\{(v_s^2 - 1)\mathcal{F}_K(v) \leq x\}],$$

where  $v = (v_0, v_1, \dots, v_K) \sim N(0, I_{K+1})$ ,  $Q_{1,K}(x)$  is the distribution function of  $Q_{1,K}$  as defined in Theorem 4.1 and  $\mathcal{F}_K(v) = \frac{v_0^2}{\sum_{j=1}^K \lambda_j v_j^2}$ . The following lemma establishes the high order expansion for Wald statistic based on the general LRV estimators when  $K$  is fixed.

**LEMMA 8.6.** *Suppose  $\sigma^2 > 0$ . Under the assumptions in Lemma 8.5 and  $H_{2,0}$ , we have*

$\sup_{x \in [0, +\infty)} |\Phi_{T,K}(x) - Q_{1,K}(x)| = O(1/T)$  and

$$\sup_{x \in [0, +\infty)} |P(F_T(K) \leq x) - \Phi_{T,K}(x)| = O(1/T^2), \quad (25)$$

with  $K$  fixed and  $T \rightarrow \infty$ .

*Proof of Lemma 8.6.* It follows directly from Lemma 8.5 that  $\sup_{x \in \mathbb{R}} |\Phi_{T,K}(x) - Q_{1,K}(x)| = O(1/T)$ . To show the second part, we first note that under the Gaussian assumption, the density function of  $\xi$  is given by  $f(u, \Sigma_\xi) = (2\pi)^{-(K+1)/2} |\Sigma_\xi|^{-1/2} \exp\left(-\frac{1}{2} u' \Sigma_\xi^{-1} u\right)$ . Taking a Taylor expansion of the density function  $f(u, \Sigma_\xi)$  around the covariance matrix  $\sigma^2 \tilde{R}$ , we get

$$f(u, \Sigma_\xi) = f(u, \sigma^2 \tilde{R}) + \frac{\partial f}{\partial \text{vec}(\Sigma)}(u, \sigma^2 \tilde{R}) \text{vec}(\Sigma_\xi - \sigma^2 \tilde{R}) + R_T(u).$$

By Lemma 8.4, the remainder term  $R_T(u)$  satisfies that  $\int_{\mathbb{R}^{K+1}} |R_T(u)| dv = O(1/T^2)$ . Following Lemma 8.1, we have  $\frac{\partial f}{\partial \text{vec}(\Sigma)}(u, \sigma^2 \tilde{R}) = f(u, \sigma^2 \tilde{R}) \left\{ \frac{1}{2\sigma^4} (\tilde{R}^{-1} u) \otimes (\tilde{R}^{-1} u) - \frac{1}{2\sigma^2} \text{vec}(\tilde{R}^{-1}) \right\}$ , which implies that

$$P(F_T(K) \leq x) = Q_{1,K}(x) \left\{ 1 - \frac{1}{2\sigma^2} \sum_{i,j=0}^K \sum_{s=0}^K L_{si} L_{sj} (\text{cov}(\xi_i, \xi_j) - \sigma^2 \tilde{R}_{ij}) \right\} + \zeta_T(x),$$

where  $\zeta_T(x) = \frac{1}{2\sigma^4} \int_{\{F_T(u;K) \leq x\}} f(u, \sigma^2 \tilde{R}) (\tilde{R}^{-1} u)' \otimes (\tilde{R}^{-1} u)' \text{vec}(\Sigma_\xi - \sigma^2 \tilde{R}) du + \int_{\{F_T(u;K) \leq x\}} R_T(u) du$  with  $F_T(u;K) = F_T(K)$ . By letting  $v = \tilde{L}u/\sigma$  and noting that  $E[\mathbf{I}\{\mathcal{F}_T(v) \leq x\} v_s v_r] = 0$  for  $s \neq r$ , we obtain

$$\begin{aligned} \zeta_T(x) &= \frac{1}{2\sigma^2} E[\mathbf{I}\{\mathcal{F}_T(v) \leq x\} (v \otimes v)'] (\tilde{L} \otimes \tilde{L}) \text{vec}(\Sigma_\xi - \sigma^2 \tilde{R}) + \int_{\{F_T(u;K) \leq x\}} R_T(u) du \\ &= \frac{1}{2\sigma^2} \sum_{i,j=0}^K \sum_{s=0}^K E[\mathbf{I}\{\mathcal{F}_T(v) \leq x\} v_s^2] L_{si} L_{sj} (\text{cov}(\xi_i, \xi_j) - \sigma^2 \tilde{R}_{ij}) + \int_{\{F_T(u;K) \leq x\}} R_T(u) du, \end{aligned}$$

where  $\mathcal{F}_K(v) = \frac{v_0^2}{\sum_{j=1}^K \lambda_j v_j^2}$  with  $v = (v_0, v_1, \dots, v_K)$  being a  $(K+1)$ -dimensional vector of iid standard normal random variables. In view of the definition of  $\Phi_{T,K}(x)$ , we get

$$\sup_{x \in \mathbb{R}} |P(F_T(K) \leq x) - \Phi_{T,K}(x)| = \sup_{x \in \mathbb{R}} \left| \int_{\{F_T(u;K) \leq x\}} R_T(u) du \right| \leq \int_{\mathbb{R}^{K+1}} |R_T(u)| du = O(1/T^2),$$

which completes the proof.  $\diamond$

**LEMMA 8.7.** Let  $\{\Sigma_{T,J+1}\} \subset \mathbb{R}^{(J+1) \times (J+1)}$  be an array of positive definite matrices with  $J+1 \leq T$ . Assume that  $\|\Sigma_{T,J+1} - \Sigma_{J+1}\|_\infty = O(J/T)$  for a sequence of positive definite matrices  $\{\Sigma_j\}_{j=1}^\infty$  with  $\sup_j \|\Sigma_j^{-1}\|_2 < \infty$ . If  $J$  satisfies that  $1/J + J^2/T \rightarrow 0$ , then we have  $\|\Sigma_{T,J+1}^{-1} - \Sigma_{J+1}^{-1}\|_\infty = O(J^2/T)$ .

*Proof.* Let  $\Sigma_{T,J+1} = \Sigma_{J+1} + R_{T,J+1}$ . For sufficiently large  $T$ , we have  $\|\Sigma_{J+1}^{-1} R_{T,J+1}\|_2 \leq (J+1) \|\Sigma_{J+1}^{-1}\|_2 \|R_{T,J+1}\|_\infty < 1$ , where we are using the fact that  $\|R_{T,J+1}\|_2 \leq (J+1) \|R_{T,J+1}\|_\infty$ . It follows that

$$\|\Sigma_{T,J+1}^{-1} - \Sigma_{J+1}^{-1}\|_\infty \leq \|\Sigma_{T,J+1}^{-1} - \Sigma_{J+1}^{-1}\|_2 \leq (J+1) \frac{\|\Sigma_{J+1}^{-1}\|_2^2 \|R_{T,J+1}\|_\infty}{1 - \|\Sigma_{J+1}^{-1} R_{T,J+1}\|_2} = O(J^2/T).$$

$\diamond$

LEMMA 8.8. Let  $\tilde{\Sigma}_{T,J+1}(y)$  be a  $(J+1) \times (J+1)$  positive definite matrix which depends on  $y \in \mathbb{R}^{J+1}$ , and  $\Sigma_{T,J+1}$  and  $\Sigma_j = \sigma^2 I_j$  satisfy the assumptions in Lemma 8.7. Assume that  $\sup_{y \in \mathbb{R}^{J+1}} \|\tilde{\Sigma}_{T,J+1}(y) - \sigma^2 I_{J+1}\|_\infty \leq \|\Sigma_{T,J+1} - \sigma^2 I_{J+1}\|_\infty = O(J/T)$ . Let  $R_{T,J+1} = \Sigma_{T,J+1} - \sigma^2 I_{J+1}$ . If  $J = o(T^{1/6})$ , we have

$$\int_{y \in \mathbb{R}^{J+1}} \left| \text{vec}(R_{T,J+1})' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \text{vec}(\Sigma)}(y, \tilde{\Sigma}_{T,J+1}(y)) \text{vec}(R_{T,J+1}) \right| dy = o(1/T).$$

*Proof.* Let  $\tilde{R}_{T,J+1}(y) = \tilde{\Sigma}_{T,J+1}(y) - \sigma^2 I_{J+1}$ . Note first that  $\sup_{y \in \mathbb{R}^{J+1}} \|\tilde{R}_{T,J+1}(y)/\sigma^2\|_2 \leq (J+1) \sup_{y \in \mathbb{R}^{J+1}} \|\tilde{R}_{T,J+1}(y)\|_\infty / \sigma^2 \leq (J+1) \|\Sigma_{T,J+1} - \sigma^2 I_{J+1}\|_\infty / \sigma^2 < 1$ , for large enough  $T$ . Following the arguments in Lemma 8.7, we know that

$$\sup_{y \in \mathbb{R}^{J+1}} \|\tilde{\Sigma}_{T,J+1}^{-1}(y) - \sigma^{-2} I_{J+1}\|_2 \leq \frac{C(J+1) \|\Sigma_{T,J+1} - \sigma^2 I_{J+1}\|_\infty}{1 - (J+1) \|\Sigma_{T,J+1} - \sigma^2 I_{J+1}\|_\infty / \sigma^2} = O(J^2/T).$$

Choose  $r = J^3/T$ . Then we have

$$\begin{aligned} y' \left( \tilde{\Sigma}_{T,J+1}^{-1}(y) - \frac{1}{(1+r)\sigma^2} I_{J+1} \right) y &= y' \left( \tilde{\Sigma}_{T,J+1}^{-1}(y) - \frac{1}{\sigma^2} I_{J+1} \right) y + \frac{r}{(r+1)\sigma^2} \|y\|^2 \\ &\geq \left( \frac{r}{(r+1)\sigma^2} - \|\tilde{\Sigma}_{T,J+1}^{-1}(y) - I_{J+1}/\sigma^2\|_2 \right) \|y\|^2 \geq 0, \end{aligned}$$

when  $T$  is sufficiently large. On the other hand, we have

$$\begin{aligned} \sup_{y \in \mathbb{R}^{J+1}} |\tilde{\Sigma}_{T,J+1}^{-1}(y)| &\leq \sup_{y \in \mathbb{R}^{J+1}} \|\tilde{\Sigma}_{T,J+1}^{-1}(y)\|_2^{J+1} \leq \left( \frac{1}{\sigma^2} + \frac{CJ^2}{T} \right)^{J+1} \\ &\leq \left| \frac{1}{(r+1)\sigma^2} I_{J+1} \right| \left( 1 + r + \frac{C(r+1)J^2\sigma^2}{T} \right)^{J+1} \\ &\leq \left| \frac{1}{(r+1)\sigma^2} I_{J+1} \right| (1 + Cr)^{(1/r)(J+1)r} \leq C \left| \frac{1}{(r+1)\sigma^2} I_{J+1} \right|. \end{aligned}$$

The above arguments imply that  $f(y, \tilde{\Sigma}_{T,J+1}(y)) \leq Cf(y, (1+r)\sigma^2 I_{J+1})$  for all  $y$ . Therefore we get

$$\begin{aligned} &\int_{y \in \mathbb{R}^{J+1}} \left| \text{vec}(R_{T,J+1})' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \text{vec}(\Sigma)}(y, \tilde{\Sigma}_{T,J+1}(y)) \text{vec}(R_{T,J+1}) \right| dy \\ &\leq C \int_{y \in \mathbb{R}^{J+1}} \left| \text{vec}(R_{T,J+1})' \{ (\tilde{\Sigma}_{T,J+1}^{-1}(y)y) \otimes (\tilde{\Sigma}_{T,J+1}^{-1}(y)y) - \text{vec}(\tilde{\Sigma}_{T,J+1}^{-1}(y)) \} \{ (\tilde{\Sigma}_{T,J+1}^{-1}(y)y) \otimes (\tilde{\Sigma}_{T,J+1}^{-1}(y)y) \right. \\ &\quad \left. - \text{vec}(\tilde{\Sigma}_{T,J+1}^{-1}(y)) \}' \text{vec}(R_{T,J+1}) \right| f(y, (1+r)\sigma^2 I_{J+1}) dy \\ &\quad + C \int_{y \in \mathbb{R}^{J+1}} \left| \text{vec}(R_{T,J+1})' \{ (\tilde{\Sigma}_{T,J+1}^{-1}(y)yy' \tilde{\Sigma}_{T,J+1}^{-1}(y)) \otimes \tilde{\Sigma}_{T,J+1}^{-1}(y) + \tilde{\Sigma}_{T,J+1}^{-1}(y) \otimes (\tilde{\Sigma}_{T,J+1}^{-1}(y)yy' \tilde{\Sigma}_{T,J+1}^{-1}(y)) \right. \\ &\quad \left. - \tilde{\Sigma}_{T,J+1}^{-1}(y) \otimes \tilde{\Sigma}_{T,J+1}^{-1}(y) \} \text{vec}(R_{T,J+1}) \right| f(y, (1+r)\sigma^2 I_{J+1}) dy \leq CJ^6/T^2 = o(1/T), \end{aligned}$$

where the first inequality in the last row comes from the fact that  $\sup_{y \in \mathbb{R}^{J+1}} \|\tilde{\Sigma}_{T,J+1}^{-1}(y) - \sigma^{-2} I_{J+1}\|_\infty \leq \sup_{y \in \mathbb{R}^{J+1}} \|\tilde{\Sigma}_{T,J+1}^{-1}(y) - \sigma^{-2} I_{J+1}\|_2 = O(J^2/T)$ .  $\diamond$

LEMMA 8.9. Recall from Section 5.2 that  $Q_{1,J}(x)$  is the distribution function of  $Q_{1,J}$  for  $J =$

$1, 2, \dots, +\infty$ . Then we have

$$\sup_{x \in [0, +\infty)} |Q_{1,J}(x) - Q_{1,\infty}(x)| = O\left(\sum_{j=J+1}^{\infty} \lambda_j\right). \quad (26)$$

*Proof.* Let  $U(J) = \sum_{j=1}^J \lambda_j v_j^2$  and  $V(J) = \sum_{j=J+1}^{\infty} \lambda_j v_j^2$ , then  $Q_{1,J} = v_0^2/U(J)$ . For any  $x \in [0, +\infty)$  and large enough  $J$  with  $J \geq 3$ , we have,

$$\begin{aligned} |Q_{1,J}(x) - Q_{1,\infty}(x)| &= |E[E[\mathbf{I}\{Q_{1,J} \leq x\}|U(J)]] - E[E[\mathbf{I}\{Q_{1,\infty} \leq x\}|U(\infty)]]| \\ &= |E[G_1(xU(J))] - E[G_1(xU(\infty))]| \\ &= |E[G_1(xU(J) + xV(J))] - E[G_1(xU(J))]| \\ &= |E[xV(J)G'_1(x\tilde{U}(J))]| = \left| E\left[\frac{V(J)}{\tilde{U}(J)}x\tilde{U}(J)G'_1(x\tilde{U}(J))\right] \right| \\ &\leq CE\left[\frac{V(J)}{\tilde{U}(J)}\right] \leq CE[V(J)]E\left[\frac{1}{U(J)}\right] \leq C\sum_{j=J+1}^{\infty} \lambda_j, \end{aligned}$$

where  $U(J) \leq \tilde{U}(J) \leq U(J) + V(J)$  and  $C$  does not depend on  $x$ . Note that we are using the mean value theorem, and the facts that  $E[1/U(J)] \leq E[1/(\lambda_3\chi_3^2)] < \infty$  and  $\sup_{x \in \mathbb{R}} |xG'_1(x)| < \infty$ .  $\diamond$

**LEMMA 8.10.** Let  $V_T(J) = \sum_{j=J+1}^{\infty} \lambda_j \xi_j^2$ . Assume that  $\sup_{1 \leq i \leq \infty} \sup_{t \in [0,1]} \phi_i(t) < \infty$  and  $\{X_t\}$  is a stationary Gaussian time series. Then we have  $EV_T^2(J) = O((\sum_{j=J+1}^{\infty} \lambda_j)^2)$ .

*Proof.* Let  $\sigma_{ij} = \gamma_X(i-j)$ . For  $i, j \geq J+1$ , we have

$$\begin{aligned} E[\xi_i^2 \xi_j^2] &= \frac{1}{T^2} \sum_{i_1, i_2=1}^T \sum_{j_1, j_2=1}^T \phi_i^0(i_1/T) \phi_i^0(i_2/T) \phi_j^0(j_1/T) \phi_j^0(j_2/T) E[(X_{i_1} - \mu)(X_{i_2} - \mu)(X_{j_1} - \mu)(X_{j_2} - \mu)] \\ &= \frac{1}{T^2} \sum_{i_1, i_2=1}^T \sum_{j_1, j_2=1}^T \phi_i^0(i_1/T) \phi_i^0(i_2/T) \phi_j^0(j_1/T) \phi_j^0(j_2/T) (\sigma_{i_1 i_2} \sigma_{j_1 j_2} + \sigma_{i_1 j_1} \sigma_{i_2 j_2} + \sigma_{i_1 j_2} \sigma_{i_2 j_1}) \\ &= I_{1,T} + I_{2,T} + I_{3,T}. \end{aligned}$$

For the first term, we have

$$I_{1,T} = \left( \frac{1}{T} \sum_{i_1, i_2=1}^T \phi_i^0(i_1/T) \phi_i^0(i_2/T) \sigma_{i_1 i_2} \right) \left( \frac{1}{T} \sum_{j_1, j_2=1}^T \phi_j^0(j_1/T) \phi_j^0(j_2/T) \sigma_{j_1 j_2} \right) = L_{1,T} L_{2,T}.$$

Note that

$$\begin{aligned} |L_{1,T}| &= \left| \frac{1}{T} \sum_{h=1-T}^{T-1} \sum_{1 \leq i_1, i_1+h \leq T} \phi_i^0(i_1/T) \phi_i^0(h/T) \gamma_X(h) \right| \leq C \sum_{h=-\infty}^{+\infty} |\gamma_X(h)| \frac{1}{T} \sum_{1 \leq i_1 \leq T} |\phi_i^0(i_1/T)| \\ &\leq C \sum_{h=-\infty}^{+\infty} |\gamma_X(h)|, \end{aligned}$$

which implies that  $|I_{1,T}| \leq C(\sum_{h=-\infty}^{+\infty} |\gamma_X(h)|)^2$ . Similar arguments apply to the other terms  $I_{2,T}$  and  $I_{3,T}$ . We then have  $\sup_{J+1 \leq i, j \leq \infty} E[\xi_i^2 \xi_j^2] < C$ . Therefore, we obtain  $E[V_T(J)^2] = \sum_{i=J+1}^{\infty} \sum_{j=J+1}^{\infty} \lambda_i \lambda_j E[\xi_i^2 \xi_j^2] \leq C(\sum_{i=J+1}^{\infty} \lambda_i)^2$ .  $\diamond$

LEMMA 8.11. Assume the eigenfunctions are continuously differentiable, mean zero and uniformly bounded, and  $\sum_{j=1}^{\infty} \lambda_j < \infty$ . Suppose that  $\{X_i\}$  is a stationary Gaussian time series with  $\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty$ . When  $1/J + J/T \rightarrow 0$ , we have

$$\sup_{x \in [0, +\infty)} |P(F_T(J) \leq x) - P(F_T(\infty) \leq x)| = O \left( \left( \sum_{j=J+1}^{\infty} \lambda_j \right)^{1/3} \right).$$

Recall that  $F_T(J) = \frac{\xi_0^2}{\sum_{j=1}^J \lambda_j \xi_j^2}$  for  $J = 1, 2, \dots, \infty$ .

Proof. Let  $R_T(J) = F_T(J) - F_T(\infty) = \frac{\xi_0^2 V_T(J)}{(\sum_{j=1}^{\infty} \lambda_j \xi_j^2)(\sum_{j=1}^J \lambda_j \xi_j^2)}$ . For any  $\delta > 0$ , we have

$$P(F_T(\infty) \leq x - \delta) - P(|R_T(J)| \geq \delta) \leq P(F_T(J) \leq x) \leq P(F_T(\infty) \leq x + \delta) + P(|R_T(J)| \geq \delta). \quad (27)$$

Observe that

$$P(|R_T(J)| \geq \delta) \leq \frac{E|R_T(J)|}{\delta} \leq \frac{(E[V_T^2(J)])^{1/2}}{\delta} \left( E \left[ \frac{\xi_0^4}{(\sum_{j=1}^J \lambda_j \xi_j^2)^4} \right] \right)^{1/2}.$$

Choose a fixed  $J_0 \geq 9$ , denote by  $\hat{\Sigma}_{T, J_0+1}$  the covariance matrix of  $(\xi_0, \dots, \xi_{J_0})$ . By Lemma 8.5, we know that  $\|\hat{\Sigma}_{T, J_0+1} - \sigma^2 I_{J_0+1}\|_2 \leq (J_0 + 1) \|\hat{\Sigma}_{T, J_0+1} - \sigma^2 I_{J_0+1}\|_{\infty} = O(1/T)$ . For large enough  $T$ , we have  $\|\hat{\Sigma}_{T, J_0+1}\|_2 \leq 2\sigma^2$ . Let  $\tilde{\lambda} = \min(1, \frac{1}{2\sigma^2}) > 0$ , we know that  $\hat{\Sigma}_{T, J_0+1}^{-1} - \tilde{\lambda} I_{J_0+1}$  is semi-positive definite, i.e., for any  $x \in \mathbb{R}^{J_0+1}$ ,  $x' \hat{\Sigma}_{T, J_0+1}^{-1} x \geq \tilde{\lambda} x' x$ . Using similar arguments in Lemma 8.3, we know that  $|\hat{\Sigma}_{T, J_0+1}|^{-1} \leq \|\hat{\Sigma}_{T, J_0+1}^{-1}\|_2^{J_0+1} \leq (2/\sigma^2)^{J_0+1}$  for large enough  $T$ . For any  $J \geq J_0$ , we have

$$\begin{aligned} E \left[ \frac{\xi_0^4}{(\sum_{j=1}^J \lambda_j \xi_j^2)^4} \right] &\leq E \left[ \frac{\xi_0^4}{\lambda_{J_0}^4 (\sum_{j=1}^{J_0} \xi_j^2)^4} \right] \\ &\leq \frac{1}{(2\pi)^{(J_0+1)/2} |\hat{\Sigma}_{T, J_0+1}|^{1/2}} \int_{w \in \mathbb{R}^{J_0+1}} \frac{w_0^4}{\lambda_{J_0}^4 (\sum_{j=1}^{J_0} w_j^2)^4} \exp(-\tilde{\lambda} w' w / 2) dw \\ &\leq C E[(\chi_1^2)^2] E[(1/\chi_{J_0}^2)^4] < \infty, \end{aligned}$$

where  $w = (w_0, w_1, \dots, w_{J_0})$  and  $\chi_m^2$  denotes a chi-square random variable with  $m$  degrees of freedom. By Lemma 8.10, we obtain

$$P(|R_T(J)| \geq \delta) \leq C \left( \sum_{j=J+1}^{\infty} \lambda_j \right) / \delta. \quad (28)$$

In what follows, we show that  $\sup_{x \in [0, \infty)} |P(F_T(\infty) \leq x \pm \delta) - P(F_T(\infty) \leq x)| \leq C\sqrt{\delta}$  for any  $\delta > 0$ . Let  $X = (X_1, X_2, \dots, X_T)'$ ,  $l_T = (1, 1, \dots, 1)'$ ,  $X^* = X - l_T \mu_0$  and  $\Omega_T = \text{cov}(X)$ . Then the GLS estimate of  $\mu$  is given by  $\hat{\mu}_{GLS} = (l_T' \Omega_T^{-1} l_T)^{-1} l_T' \Omega_T^{-1} X$  and  $\hat{\mu}_{OLS} - \mu_0 = \hat{\mu}_{GLS} - \mu_0 + \frac{1}{T} l_T' \tilde{X}$ , where  $\tilde{X} = (I_T - l_T (l_T' \Omega_T^{-1} l_T)^{-1} l_T' \Omega_T^{-1}) X^*$ . The following facts which can be found in Sun et al.(2008) play an important role in the proof presented below: (1)  $\hat{\mu}_{GLS} - \mu_0$  is independent of  $\tilde{X}$ ; (2)  $\hat{\mu}_{GLS} - \mu_0$  is independent of  $X - l_T \hat{\mu}_{OLS}$ . Notice that  $\hat{D}_T = \sum_{j=1}^{\infty} \lambda_j \xi_j^2 = \frac{1}{T} (X - l_T \hat{\mu}_{OLS})' \mathcal{G}_T (X - l_T \hat{\mu}_{OLS})$  with  $\mathcal{G}_T = (\mathcal{G}(i/T, j/T))_{i,j=1}^T$ . Then  $\hat{\mu}_{GLS} - \mu_0$  is also independent of  $\hat{D}_T$ . Define  $\sigma_{GLS}^2 = T \text{var}(\hat{\mu}_{GLS}) = T(l_T' \Omega_T^{-1} l_T)^{-1}$ . Denote by  $\Phi_{norm}$  and  $\phi_{norm}$  the cumulative distribution function and density function



of the standard normal distribution. Therefore, we get

$$\begin{aligned}
P(F_T(\infty) \leq x) &= 2P\left(\frac{\sqrt{T}(\hat{\mu}_{OLS} - \mu_0)}{\sqrt{\hat{D}_T}} \leq \sqrt{x}\right) - 1 = 2P\left(\sqrt{T}(\hat{\mu}_{OLS} - \mu_0) \leq \sqrt{x\hat{D}_T}\right) - 1 \\
&= 2P\left(\sqrt{T}(\hat{\mu}_{GLS} - \mu_0)/\sigma_{GLS} \leq \sqrt{x\hat{D}_T}/\sigma_{GLS} - l'_T\tilde{X}/(\sqrt{T}\sigma_{GLS})\right) - 1 \\
&= 2E\left[\Phi_{norm}\left(\sqrt{x\hat{D}_T}/\sigma_{GLS} - l'_T\tilde{X}/(\sqrt{T}\sigma_{GLS})\right)\right] - 1,
\end{aligned}$$

which implies that for  $x, \delta \geq 0$  with  $x - \delta \geq 0$ ,

$$\begin{aligned}
&|P(F_T(\infty) \leq x \pm \delta) - P(F_T(\infty) \leq x)| \\
&= \left|2E\left[\Phi_{norm}\left(\sqrt{(x \pm \delta)\hat{D}_T}/\sigma_{GLS} - l'_T\tilde{X}/(\sqrt{T}\sigma_{GLS})\right)\right] - 2E\left[\Phi_{norm}\left(\sqrt{x\hat{D}_T}/\sigma_{GLS} - l'_T\tilde{X}/(\sqrt{T}\sigma_{GLS})\right)\right]\right| \\
&= 2\left|(\sqrt{x \pm \delta} - \sqrt{x})E\left[\sqrt{\hat{D}_T}\phi_{norm}(a^* - l'_T\tilde{X}/(\sqrt{T}\sigma_{GLS}))/\sigma_{GLS}\right]\right| \\
&\leq C\sqrt{\delta}E[\sqrt{\hat{D}_T}/\sigma_{GLS}] < C\sqrt{\delta}(E[\hat{D}_T])^{1/2}/\sigma_{GLS} < C\sqrt{\delta},
\end{aligned} \tag{29}$$

with  $\sqrt{x\hat{D}_T}/\sigma_{GLS} \leq a^* \leq \sqrt{(x + \delta)\hat{D}_T}/\sigma_{GLS}$  or  $\sqrt{(x - \delta)\hat{D}_T}/\sigma_{GLS} \leq a^* \leq \sqrt{x\hat{D}_T}/\sigma_{GLS}$ . Here we are using the fact that  $\sigma_{GLS}^2 = \sigma^2 + O(1/T)$  and  $E[\hat{D}_T]$  is uniformly bounded for all  $T$ . Choosing  $\delta = (\sum_{j=J+1}^{\infty} \lambda_j)^{2/3}$ , the conclusion follows in view of (27), (28) and (29).  $\diamond$

**LEMMA 8.12.** *Under the assumptions in Theorem 5.2, we have  $\|\Sigma_{\xi, J+1} - \sigma^2 I_{J+1}\|_{\infty} = O(J/T)$  with  $J \leq T$ , where  $\Sigma_{\xi, J+1}$  denotes the covariance matrix of  $(\xi_0, \xi_1, \dots, \xi_J)$ .*

*Proof of Lemma 8.12.* Using the arguments in Lemma 8.5, we have for any  $1 \leq s \leq J$ ,

$$|\text{cov}(\xi_0, \xi_s)| \leq C \left| \frac{1}{T^2} \sum_{i=1}^T \phi_s(i/T) \right| + \frac{1}{T} \sum_{0 < h < T} \left| \gamma_X(h) \left\{ \sum_{i=1}^h \phi_s(i/T) + \sum_{i=T-h+1}^T \phi_s(i/T) \right\} \right| \leq C/T,$$

where  $C$  is a generic constant which does not depend on  $s$ . Again by the arguments in Lemma 8.5, we have

$$\begin{aligned}
|\text{cov}(\xi_s, \xi_r) - \sigma^2 \delta_{sr}| &\leq \sum_{h=1-T}^{T-1} |\gamma_X(h) C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t))| + \sum_{|h| \geq T} |\gamma_X(h)| \delta_{sr} \\
&\quad + \frac{1}{T} \sum_{h=1}^{T-1} \left| \gamma_X(h) \left\{ \sum_{j=1}^h \phi_r^0\left(\frac{j}{T}\right) \phi_s^0\left(\frac{j}{T}\right) + \sum_{j=T-h+1}^T \phi_r^0\left(\frac{j}{T}\right) \phi_s^0\left(\frac{j}{T}\right) \right\} \right| \\
&\quad + |J_{1,T}| + |J_{2,T}|, \quad 1 \leq s, r \leq J,
\end{aligned}$$

where  $J_{1,T}$ ,  $J_{2,T}$  and  $C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t))$  are defined in the proof of Lemma 8.5. By the Trapezoidal rule and the assumption that  $\sup_{1 \leq i \leq J} \sup_{t \in [0,1]} |\phi_i''(t)| < CJ^2$ , we have

$$|C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t))| \leq C(J^2/T^2 + 1/T), \tag{30}$$

which implies that  $|\text{cov}(\xi_s, \xi_r) - \sigma^2 \delta_{sr}| \leq CJ/T + |J_{1,T}| + |J_{2,T}|$  for  $J \leq T$ . By the mean value theorem

and the assumption that  $\sup_{1 \leq i \leq J} \sup_{t \in [0,1]} |\phi'_i(t)| < CJ$ , we get

$$\begin{aligned} |J_{1,T}| &\leq \frac{1}{T} \sum_{h=1}^{T-1} |\gamma_X(h)| \left| \sum_{j=1}^{T-h} \phi_r^0\left(\frac{j}{T}\right) \left\{ \phi_s^0\left(\frac{j+h}{T}\right) - \phi_s^0\left(\frac{j}{T}\right) \right\} \right| \\ &\leq \frac{CJ}{T} \sum_{h=1}^{T-1} |h\gamma_X(h)| \frac{1}{T} \sum_{j=1}^{T-h} \left| \phi_r^0\left(\frac{j}{T}\right) \right| \leq \frac{CJ}{T}. \end{aligned} \quad (31)$$

Using the same argument for  $J_{2,T}$ , we get  $|\text{cov}(\xi_s, \xi_r) - \sigma^2 \delta_{sr}| \leq CJ/T$ , which completes the proof.  $\diamond$

*Proof of Theorem 5.2.* Suppose  $J = o(T^{1/6})$ . By Lemma 8.12, we know  $\|\Sigma_{\xi, J+1} - \sigma^2 I_{J+1}\|_\infty = O(J/T)$ . Using Lemma 8.8 and similar arguments in the proof of Lemma 8.6, we can show that

$$\sup_{x \in \mathbb{R}} |P(F_T(J) \leq x) - Q_{1,J}(x) - \psi_{J,T}(x)| = o(1/T),$$

where  $\psi_{J,T}(x) = \frac{1}{2\sigma^2} \sum_{i=0}^J (\text{var}(\xi_i) - \sigma^2) E[(v_i^2 - 1) \mathbf{I}\{\mathcal{F}_J(v) \leq x\}]$  with  $v = (v_0, v_1, \dots, v_J) \sim N(0, I_{J+1})$ . Next, we show that  $\psi_J(x)$  converges uniformly as  $J \rightarrow +\infty$ . Note first that

$$\begin{aligned} \sup_{x \in [0, +\infty)} |\psi_{J+p,T}(x) - \psi_{J,T}(x)| &\leq \sup_{x \in [0, +\infty)} \left| \frac{1}{2\sigma^2} \sum_{i=J+1}^{J+p} (\text{var}(\xi_i) - \sigma^2) E[(v_i^2 - 1) \mathbf{I}\{\mathcal{F}_{J+p}(v) \leq x\}] \right| \\ &+ \sup_{x \in [0, +\infty)} \left| \frac{1}{2\sigma^2} \sum_{i=1}^J (\text{var}(\xi_i) - \sigma^2) E[(v_i^2 - 1) (\mathbf{I}\{\mathcal{F}_{J+p}(v) \leq x\} - \mathbf{I}\{\mathcal{F}_J(v) \leq x\})] \right| \\ &+ \sup_{x \in [0, +\infty)} \left| \frac{1}{2\sigma^2} (\text{var}(\xi_0) - \sigma^2) E[(v_0^2 - 1) (\mathbf{I}\{\mathcal{F}_{J+p}(v) \leq x\} - \mathbf{I}\{\mathcal{F}_J(v) \leq x\})] \right| \\ &= I_1 + I_2 + I_3, \end{aligned}$$

for any  $J, p \in \mathbb{Z}^+$ . In view of (30) and (31), we have

$$|\text{var}(\xi_i) - \sigma^2| < C(i/T + i^2/T^2), \quad (32)$$

for  $1 \leq i < \infty$ . Hence we get, for sufficiently large  $J$ ,

$$\begin{aligned}
I_1 &\leq \frac{1}{2\sigma^2} \sup_{x \in [0, +\infty)} \sum_{i=J+1}^{J+p} \left| (\text{var}(\xi_i) - \sigma^2) E \left[ (v_i^2 - 1) G_1 \left( x \sum_{j=1}^{J+p} \lambda_j v_j^2 \right) \right] \right| \\
&\leq \frac{C}{T} \sup_{x \in [0, +\infty)} \sum_{i=J+1}^{J+p} (i + i^2/T) \left| E \left[ (v_i^2 - 1) G_1 \left( x \sum_{j=1}^{J+p} \lambda_j v_j^2 \right) \right] \right| \\
&\leq \frac{C}{T} \sup_{x \in [0, +\infty)} \sum_{i=J+1}^{J+p} (i + i^2/T) \left| E \left[ (v_i^2 - 1) \left\{ G_1 \left( x \sum_{j \neq i} \lambda_j v_j^2 \right) + \lambda_i v_i^2 x G_1'(y_i) \right\} \right] \right| \\
&\leq \frac{C}{T} \sup_{x \in [0, +\infty)} \sum_{i=J+1}^{J+p} (i + i^2/T) \lambda_i E \left[ v_i^2 (v_i^2 + 1) x G_1'(y_i) \right] \leq \frac{C}{T} \sum_{i=J+1}^{J+p} (i + i^2/T) \lambda_i E \left[ \frac{v_i^2 (v_i^2 + 1)}{\sum_{j \neq i} \lambda_j v_j^2 + \alpha_i \lambda_i^2 v_i^2} \right] \\
&\leq \frac{C}{T} \sum_{i=J+1}^{J+p} (i + i^2/T) \lambda_i E \left[ v_i^2 (v_i^2 + 1) \right] E \left[ \frac{1}{\sum_{j \neq i} \lambda_j v_j^2} \right] \\
&\leq \frac{C}{T} \left\{ \sum_{i=J+1}^{+\infty} i \lambda_i + \frac{1}{T} \sum_{i=J+1}^{+\infty} i^2 \lambda_i \right\} = O \left( \frac{J^{-a+2}}{T} \right),
\end{aligned} \tag{33}$$

where  $y_i = x(\sum_{j \neq i} \lambda_j v_j^2 + \alpha_i \lambda_i^2 v_i^2)$  for some  $0 \leq \alpha_i \leq 1$ . On the other hand, we get

$$\begin{aligned}
I_2 &\leq \frac{CJ}{T} \sup_{x \in [0, +\infty)} \sum_{i=1}^J \left| E \left[ (v_i^2 - 1) \left\{ G_1 \left( x \sum_{j=1}^{J+p} \lambda_j v_j^2 \right) - G_1 \left( x \sum_{j=1}^J \lambda_j v_j^2 \right) \right\} \right] \right| \\
&\leq \frac{CJ}{T} \sup_{x \in [0, +\infty)} \sum_{i=1}^J \left| E \left[ x(v_i^2 - 1) \left( \sum_{j=J+1}^{J+p} \lambda_j v_j^2 \right) \left\{ G_1' \left( x \sum_{j=1}^J \lambda_j v_j^2 + x \alpha \sum_{j=J+1}^{J+p} \lambda_j v_j^2 \right) \right\} \right] \right| \\
&\leq \frac{CJ}{T} \left( \sum_{j=J+1}^{J+p} \lambda_j \right) E \left[ \frac{\sum_{i=1}^J (v_i^2 + 1)}{\sum_{j=1}^J \lambda_j v_j^2} \right] \leq \frac{CJ^2}{T} \left( \sum_{j=J+1}^{\infty} \lambda_j \right) = O \left( \frac{J^{-a+3}}{T} \right).
\end{aligned}$$

Finally using the Cauchy-Schwarz inequality and similar arguments in Lemma 8.9, we know

$$\begin{aligned}
I_3 &\leq \frac{C}{T} \{E[(v_0^2 - 1)^2]\}^{1/2} \sup_{x \in [0, +\infty)} \{E[(\mathbf{I}\{\mathcal{F}_{J+p}(v) \leq x\} - \mathbf{I}\{\mathcal{F}_J(v) \leq x\})^2]\}^{1/2} \\
&\leq \frac{C}{T} \sup_{x \in [0, +\infty)} \{E[|\mathbf{I}\{\mathcal{F}_{J+p}(v) \leq x\} - \mathbf{I}\{\mathcal{F}_J(v) \leq x\}|]\}^{1/2} \leq \frac{C}{T} \left( \sum_{j=J}^{+\infty} \lambda_j \right)^{1/2} = O \left( \frac{J^{(-a+1)/2}}{T} \right).
\end{aligned}$$

Therefore, it is straightforward to see that  $\sup_{x \in [0, \infty)} |\psi_{J,T}(x) - \psi_T(x)| = O(J^{(-a+1)/2}/T)$  and  $\sup_{x \in [0, \infty)} |\psi_T(x)| = O(1/T)$ , which imply that

$$\sup_{x \in [0, \infty)} |P(F_T(J) \leq x) - Q_{1,J}(x) - \psi_T(x)| = o(1/T),$$

for  $J = o(T^{1/6})$ . Let  $J = T^{1/6}/\log(T)$  and note that  $(\sum_{j=J+1}^{\infty} \lambda_j)^{1/3} = o(1/T)$ . The proof is completed in view of Lemma 8.9 and Lemma 8.11.  $\diamond$

*Proof of Proposition 5.3.* Under the assumption that  $\sup_{x \in \mathbb{R}} |\mathcal{K}(x)| \leq 1$  and  $\int_{-\infty}^{+\infty} |\mathcal{K}(x)| dx < \infty$ , we

have

$$\begin{aligned} \sum_{j=1}^{+\infty} (\tilde{\lambda}_{j,b})^2 &= \int_0^1 \int_0^1 \tilde{\mathcal{G}}_b^2(r, t) dr dt \leq \sup_{t \in [0,1]} \int_0^1 \tilde{\mathcal{G}}_b^2(r, t) dr \leq 4 \sup_{t \in [0,1]} \int_0^1 |\tilde{\mathcal{G}}_b(r, t)| dr \\ &\leq 16 \sup_{t \in [0,1]} \int_{-t}^{1-t} |\mathcal{K}_b(r)| dr \leq 16 \int_{-\infty}^{+\infty} |\mathcal{K}_b(r)| dr \leq Cb, \end{aligned}$$

and  $\tilde{\lambda}_{1,b} \leq (\int_0^1 \int_0^1 \tilde{\mathcal{G}}_b^2(r, t) dr dt)^{1/2} \leq C\sqrt{b}$ . Suppose  $\{\tilde{a}_i\}$  is a sequence of random variables such that  $0 \leq \tilde{a}_i \leq 1$ . Using the fact that  $\sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b} = \int_0^1 \tilde{\mathcal{G}}_b(r, r) dr = 1 + O(b)$ , we get

$$\sup_i E \left( \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 + \tilde{a}_i \tilde{\lambda}_{i,b} v_i^2 - 1 \right)^2 = \sup_i \left\{ 2 \sum_{j \neq i}^{+\infty} (\tilde{\lambda}_{j,b})^2 + (\tilde{\lambda}_{i,b})^2 E(\tilde{a}_i v_i^2 - 1)^2 \right\} + O(b) \leq Cb. \quad (34)$$

By the Talyor expansion, we have

$$\begin{aligned} \psi_{T,b}(x) &= \frac{1}{2\sigma^2} \sum_{i=1}^{+\infty} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E[(v_i^2 - 1) \mathbf{I}\{\mathcal{F}_\infty(v) \leq x\}] + O(1/T) \\ &= \frac{1}{2\sigma^2} \sum_{i=1}^{+\infty} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E \left[ (v_i^2 - 1) G_1 \left( x \sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 \right) \right] + O(1/T) \\ &= \frac{x}{\sigma^2} \sum_{i=1}^{+\infty} \tilde{\lambda}_{i,b} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E \left[ G_1' \left( x \left( \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 \right) \right) \right] \\ &\quad + \frac{x^2}{4\sigma^2} \sum_{i=1}^{+\infty} (\tilde{\lambda}_{i,b})^2 (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E \left[ v_i^4 (v_i^2 - 1) G_1'' \left( x \left( \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 + a_i \tilde{\lambda}_{i,b} v_i^2 \right) \right) \right] + O(1/T) \\ &= I_{1T,b} + I_{2T,b} + O(1/T), \end{aligned}$$

where  $0 \leq a_i \leq 1$ . Let  $A_{i,b} = E \left[ G_1' \left( x \left( \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 \right) \right) \right]$ ,  $B_{i,b} = \tilde{\lambda}_{i,b} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2)$ ,  $C_{i,b} = \sum_{j=1}^i B_{j,b}$  and  $S_{N,b} = \sum_{i=1}^N A_{i,b} B_{i,b}$ . Using summation by parts, we have  $S_{N,b} = A_{N,b} C_{N,b} - \sum_{i=1}^{N-1} (A_{i+1,b} - A_{i,b}) C_{i,b}$ . Note that  $\{A_{i,b}\}_{i=1}^{+\infty}$  is a nonincreasing sequence and  $\lim_{b \rightarrow 0} \sup_i A_{i,b} = G_1'(x)$  as seen from (34). Let  $\hat{D}_{T,b}$  be defined by replacing  $\phi_j$  and  $\lambda_j$  with  $\tilde{\phi}_{j,b}$  and  $\tilde{\lambda}_{j,b}$  in the definition of  $\hat{D}_T$ . It is not hard to see that as  $b + 1/(bT) \rightarrow 0$ ,

$$\lim_{N \rightarrow +\infty} A_{N,b} C_{N,b} = \sigma^2 G'(x) \left( E[\hat{D}_{T,b}]/\sigma^2 - \sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b} \right) (1 + o(1)) = - \frac{G'(x) g_q \sum_{h=-\infty}^{+\infty} |h|^q \gamma_X(h)}{(bT)^q} (1 + o(1)) + O(1/T),$$

where we have used the fact  $E[\hat{D}_{T,b}]/\sigma^2 - \sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b} = - \frac{g_q \sum_{h=-\infty}^{+\infty} |h|^q \gamma_X(h)}{\sigma^2 (bT)^q} (1 + o(1)) + O(1/T)$ , which can be proved by using similar arguments in the proof of Lemma 2 in Sun et al. (2008). On the other hand, observe that  $|\sum_{i=1}^{N-1} (A_{i+1,b} - A_{i,b}) C_{i,b}| \leq \sup_{i \in \mathbb{N}} |C_{i,b}| \sum_{i=1}^{N-1} (A_{i,b} - A_{i+1,b}) \leq \sup_{i \in \mathbb{N}} |C_{i,b}| (A_{1,b} - \lim_{N \rightarrow +\infty} A_{N,b}) = o(|\lim_{N \rightarrow +\infty} C_{N,b}|)$  as  $b + 1/(bT) \rightarrow 0$ , for all  $N$ . Hence we get

$$I_{1T,b} = - \frac{x G'(x) g_q \sum_{h=-\infty}^{+\infty} |h|^q \gamma_X(h)}{\sigma^2 (bT)^q} (1 + o(1)) + O(1/T).$$

Define  $H_{i,b} = \tilde{\lambda}_{i,b} E \left[ v_i^4 (v_i^2 - 1) G_1'' \left( x \left( \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 + a_i \tilde{\lambda}_{i,b} v_i^2 \right) \right) \right]$  and  $\tilde{S}_{N,b} = \sum_{i=1}^{+\infty} H_{i,b} B_{i,b}$ . Again using

summation by parts, we obtain  $\tilde{S}_{N,b} = H_{N,b}C_{N,b} - \sum_{i=1}^{N-1} (H_{i+1,b} - H_{i,b})C_{i,b}$ . By (34), we can show that  $\sup_i |H_{i,b}/\tilde{\lambda}_{i,b} - 12G_1''(x)| = O(\sqrt{b})$ . Therefore, we get  $\lim_{N \rightarrow +\infty} C_{N,b}H_{N,b} = o(\lim_{N \rightarrow +\infty} C_{N,b})$  and

$$\left| \sum_{i=1}^{N-1} (H_{i+1,b} - H_{i,b})C_{i,b} \right| \leq \sup_{i \in \mathbb{N}} |C_{i,b}| \left\{ \sum_{i=1}^{N-1} (|H_{i+1,b} - 12\tilde{\lambda}_{i+1,b}G_1''(x)| + 12G''(x)(\tilde{\lambda}_{i,b} - \tilde{\lambda}_{i+1,b}) + |12\tilde{\lambda}_{i,b}G_1''(x) - H_{i,b}|) \right\} = O\left(\sqrt{b} \lim_{N \rightarrow +\infty} C_{N,b}\right).$$

The conclusion follows from the above arguments by noting that  $I_{2T,b} = o(I_{1T,b})$ .  $\diamond$

## 8.4 Proof of the main results in section 6

*Proof of Theorem 6.1.* The proof is similar to those of Lemma 8.13 and Theorem 6.2. The details are omitted.  $\diamond$

**LEMMA 8.13.** *Let  $\omega_l(x) = (1 - |x/l|)\mathbf{I}\{|x/l| < 1\}$ . Suppose that  $m^3/l^2 + (ml)^3/T + 1/m \rightarrow 0$  and  $\sum_{h=-\infty}^{+\infty} h^2|\gamma_X(h)| < \infty$ . Then under the Gaussian assumption, we have*

$$\sup_{0 \leq k \leq m} \left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\omega_l(h)\hat{\gamma}_X(h) - \sum_{h=1-T}^{T-1} g_{k,T}(h)\gamma_X(h) \right| = O_p(\sqrt{m^3/l^2 + (ml)^3/T}),$$

where  $|g_{k,T}(h)| \leq C(k|h| + |h| + 1)$  for  $0 \leq k \leq m$  and  $|h| \leq T$ , and the constant  $C$  does not depend on  $k$  and  $h$ .

*Proof of Lemma 8.13.* Note first that for any  $\epsilon > 0$ ,

$$\begin{aligned} & P\left(\sup_{0 \leq k \leq m} \left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\omega_l(h)\hat{\gamma}_X(h) - \sum_{h=1-T}^{T-1} g_{k,T}(h)\gamma_X(h) \right| > \epsilon\right) \\ & \leq \sum_{k=0}^m P\left(\left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\omega_l(h)\hat{\gamma}_X(h) - \sum_{h=1-T}^{T-1} g_{k,T}(h)\gamma_X(h) \right| > \epsilon\right) \\ & \leq \frac{1}{\epsilon^2} \sum_{k=0}^m E \left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\omega_l(h)\hat{\gamma}_X(h) - \sum_{h=1-T}^{T-1} g_{k,T}(h)\gamma_X(h) \right|^2 \\ & \leq \frac{2}{\epsilon^2} \sum_{k=0}^m E \left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\{\omega_l(h)\hat{\gamma}_X(h) - \gamma_X(h)\} \right|^2 + \frac{Cm^3}{l^2\epsilon^2}. \end{aligned}$$

Let  $z_i = X_i - E[X_i]$  and  $w_{i|h} = z_i z_{i+|h|} - \gamma_X(h)$ . Simple calculation yields that

$$\begin{aligned} & \left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\{\omega_l(h)\hat{\gamma}_X(h) - \gamma_X(h)\} \right| = \left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\omega_l(h) \{\hat{\gamma}_X(h) - \gamma_X(h)\} \right| + C(k+1)/l \\ & \leq \left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\omega_l(h) \left\{ \frac{1}{T} \sum_{i=1}^{T-|h|} w_{i|h} \right\} \right| + \left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\omega_l(h) \left\{ \frac{T-|h|}{T} \bar{z}^2 \right\} \right| \\ & \quad + \left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\omega_l(h) \left\{ \frac{\bar{z}}{T} \sum_{i=1}^{T-|h|} (z_i + z_{i+|h|}) \right\} \right| + C(k+1)/l := I_{1T} + I_{2T} + I_{3T} + C(k+1)/l, \end{aligned}$$

which implies that  $E \left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\{\omega_l(h)\hat{\gamma}_X(h) - \gamma_X(h)\} \right|^2 \leq C(EI_{1T}^2 + EI_{2T}^2 + EI_{3T}^2 + (k+1)^2/l^2)$ .

We proceed to derive the order of  $EI_{1T}^2$ . Notice that

$$\begin{aligned}
EI_{1T}^2 &= \frac{1}{T^2} \sum_{h_1, h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} \text{cov}(w_{i_1|h_1|}, w_{i_2|h_2|}) g_{k,T}(h_1) g_{k,T}(h_2) \omega_l(h_1) \omega_l(h_2) \\
&\leq \frac{C(k+1)^2}{T^2} \sum_{h_1, h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} (|h_1|+1)(|h_2|+1) |\gamma_X(i_1-i_2) \gamma_X(i_1-i_2+|h_1|-|h_2|)| \\
&\quad + \frac{C(k+1)^2}{T^2} \sum_{h_1, h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} (|h_1|+1)(|h_2|+1) |\gamma_X(i_1-i_2-|h_2|) \gamma_X(i_1-i_2+|h_1|)| \\
&\leq \frac{C(k+1)^2}{T^2} \sum_{h_1, h_2=1-l}^{l-1} \sum_{s=1-T}^{T-1} (T-|s|)(|h_1|+1)(|h_2|+1) |\gamma_X(s) \gamma_X(s+|h_1|-|h_2|)| \\
&\quad + \frac{C(k+1)^2}{T^2} \sum_{h_1, h_2=1-l}^{l-1} \sum_{s=1-T}^{T-1} (T-|s|)(|h_1|+1)(|h_2|+1) |\gamma_X(s-|h_2|) \gamma_X(s+|h_1|)| := \mathcal{J}_{1,T} + \mathcal{J}_{2,T}.
\end{aligned}$$

Then we get

$$\begin{aligned}
\mathcal{J}_{1,T} &\leq \frac{C(k+1)^2}{T} \sum_{h_1, h_2=1-l}^{l-1} (|h_1|+1)(|h_2|+1) \sum_{s=-\infty}^{+\infty} |\gamma_X(s) \gamma_X(s+|h_1|-|h_2|)| \\
&\leq \frac{C(k+1)^2 l^2}{T} \sum_{s=-\infty}^{+\infty} |\gamma_X(s)| \sum_{h_1, h_2=1-l}^{l-1} |\gamma_X(s+|h_1|-|h_2|)| \leq \frac{C(k+1)^2 l^3}{T},
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{J}_{2,T} &\leq \frac{C(k+1)^2}{T} \sum_{h_1, h_2=1-l}^{l-1} (|h_1|+1)(|h_2|+1) \sum_{s=-\infty}^{+\infty} |\gamma_X(s) \gamma_X(s+|h_1|+|h_2|)| \\
&\leq \frac{C(k+1)^2 l}{T} \sum_{s=-\infty}^{+\infty} |\gamma_X(s)| \sum_{h_1, h_2=1-l}^{l-1} (|h_1|+|h_2|+1) |\gamma_X(s+|h_1|+|h_2|)| \\
&\leq \frac{C(k+1)^2 l}{T} \sum_{s=-\infty}^{+\infty} |\gamma_X(s)| \sum_{v=1}^{2l-1} v^2 |\gamma_X(s+v)| \leq \frac{C(k+1)^2 l}{T}.
\end{aligned}$$

It implies that  $EI_{1T}^2 \leq \frac{C(k+1)^2 l^3}{T}$ . Applying similar arguments to  $I_{2T}$  and  $I_{3T}$ , we get  $EI_{2T}^2 \leq C(k+1)^2 l^4/T^2$  and  $EI_{3T}^2 \leq C(k+1)^2 l^4/T^2$ . Note the constant  $C$  above does not depend on  $m$  by the assumption. We then have

$$P \left( \sup_{0 \leq k \leq m} \left| \sum_{h=1-l}^{l-1} g_{k,T}(h) \omega_l(h) \hat{\gamma}_X(h) - \sum_{h=1-T}^{T-1} g_{k,T}(h) \gamma_X(h) \right| > \epsilon \right) \leq \frac{C}{\epsilon^2} (m^3/l^2 + (ml)^3/T) \rightarrow 0.$$

◇

*Proof of Theorem 6.2.* We choose  $m$  so that  $m^3/l^2 + (ml)^3/T + 1/m \rightarrow 0$  (e.g.,  $l \asymp T^{1/5}$  and  $m \asymp T^{2/15-\epsilon}$  for some  $\epsilon > 0$ ). From equation (24) in Lemma 8.5, we know that

$$\text{var}(\xi_i) - \sigma^2 - (\text{var}^*(\xi_i^*) - \hat{\sigma}^2) = \frac{1}{T} \left\{ \sum_{h=1-T}^{T-1} g_{i,T}(h) \gamma_X(h) - \sum_{h=1-l}^{l-1} g_{i,T}(h) \omega_l(h) \hat{\gamma}_X(h) \right\} - \sum_{|h| \geq T} \gamma_X(h),$$

where  $\hat{\sigma}^2 = \sum_{h=1-l}^{l-1} \omega_l(h) \hat{\gamma}_X(h)$  and  $g_{0,T}(h) = -|h|$ ,

$$\begin{aligned} g_{i,T}(h) = & TC_T(\phi_i(s), \phi_i(t)) - \left[ \sum_{j=1}^h \left\{ \phi_i^0 \left( \frac{j}{T} \right) \right\}^2 + \sum_{j=T-h+1}^T \left\{ \phi_i^0 \left( \frac{j}{T} \right) \right\}^2 \right] \mathbf{I}\{h \geq 1\} \\ & + \left[ \sum_{j=1}^{T-h} \phi_i^0 \left( \frac{j}{T} \right) \left\{ \phi_i^0 \left( \frac{j+h}{T} \right) - \phi_i^0 \left( \frac{j}{T} \right) \right\} \right] \mathbf{I}\{h \geq 1\} \\ & + \left[ \sum_{j=1+|h|}^T \phi_i^0 \left( \frac{j}{T} \right) \left\{ \phi_i^0 \left( \frac{j+h}{T} \right) - \phi_i^0 \left( \frac{j}{T} \right) \right\} \right] \mathbf{I}\{h \leq -1\}, \end{aligned}$$

for  $1 \leq i \leq m$ . Note that  $\sup_{1 \leq i \leq m} |TC_T(\phi_i(s), \phi_i(t))| \leq C$ . It is not hard to see that  $|g_{i,T}(h)| \leq C(|ih| + |h| + 1)$  for  $0 \leq i \leq m$ . By Lemma 8.13, we know

$$\sup_{0 \leq i \leq m} |\text{var}(\xi_i) - \sigma^2 - \text{var}^*(\xi_i^*) + \hat{\sigma}^2| = O_p \left( \frac{\sqrt{m^3/l^2 + (ml)^3/T}}{T} \right).$$

Since the bootstrap sample is normally distributed and  $\sum_{h=l-l}^{l-1} h^2 \omega_l(h) |\hat{\gamma}_X(h)|$  is bounded in probability in view of the fact that  $\sum_{h=-\infty}^{+\infty} h^2 \omega_l(h) E|\hat{\gamma}_X(h)| < \infty$ , Theorem 5.2 is also applicable to the bootstrap sample, i.e.,

$$\sup_{x \in [0, \infty)} |P(F_T^*(\infty) \leq x) - Q_{1,\infty}(x) - \psi_T^*(x)| = o_p(1/T),$$

where  $\psi_T^*(x) = \frac{1}{2\hat{\sigma}^2} \sum_{i=0}^{\infty} (\text{var}^*(\xi_i^*) - \hat{\sigma}^2) E[(v_i^2 - 1) \mathbf{I}\{\mathcal{F}_{\infty}(v) \leq x\}]$ . It is not hard to show that  $\hat{\sigma}^2 - \sigma^2 = O_p(\sqrt{l/T + 1/l^2})$ . Note that  $\text{var}^*(\xi_i^*) - \hat{\sigma}^2 = \frac{1}{T} \sum_{h=1-l}^{l-1} g_{i,T}(h) \omega_l(h) \hat{\gamma}_X(h)$ , which implies that  $\sup_{1 \leq i < +\infty} \frac{|\text{var}^*(\xi_i^*) - \hat{\sigma}^2|}{i/T + i^2/T^2} = O_p(1)$  [see e.g., (32)]. Using the arguments in (33), we can show that

$$\sup_{x \in [0, +\infty)} \left| \frac{1}{2\sigma^2} \sum_{i=m+1}^{\infty} (\text{var}(\xi_i) - \text{var}^*(\xi_i^*) + \hat{\sigma}^2 - \sigma^2) E[(v_i^2 - 1) \mathbf{I}\{\mathcal{F}_{\infty}(v) \leq x\}] \right| = O_p \left( \frac{1}{Tm^{a-2}} \right).$$

Thus we get

$$\begin{aligned} \sup_{x \in [0, +\infty)} |\psi_T(x) - \psi_T^*(x)| &\leq \sup_{x \in [0, +\infty)} \left| \frac{1}{2\sigma^2} \sum_{i=0}^{\infty} (\text{var}(\xi_i) - \text{var}^*(\xi_i^*) + \hat{\sigma}^2 - \sigma^2) E[(v_i^2 - 1) \mathbf{I}\{\mathcal{F}_{\infty}(v) \leq x\}] \right| \\ &+ \sup_{x \in [0, +\infty)} \left| \left( \frac{1}{2\hat{\sigma}^2} - \frac{1}{2\sigma^2} \right) \sum_{i=1}^{\infty} (\text{var}^*(\xi_i^*) - \hat{\sigma}^2) E[(v_i^2 - 1) \mathbf{I}\{\mathcal{F}_{\infty}(v) \leq x\}] \right| \\ &\leq \frac{1}{2\sigma^2} \sup_{1 \leq i \leq m} |\text{var}(\xi_i) - \sigma^2 - \text{var}^*(\xi_i^*) + \hat{\sigma}^2| \sup_{x \in [0, +\infty)} \left| \sum_{i=1}^m E[(v_i^2 - 1) \mathbf{I}\{\mathcal{F}_{\infty}(v) \leq x\}] \right| \\ &+ \sup_{x \in [0, +\infty)} \left| \frac{1}{2\sigma^2} \sum_{i=m+1}^{\infty} (\text{var}(\xi_i) - \text{var}^*(\xi_i^*) + \hat{\sigma}^2 - \sigma^2) E[(v_i^2 - 1) \mathbf{I}\{\mathcal{F}_{\infty}(v) \leq x\}] \right| + O_p \left( \frac{\sqrt{l/T + 1/l^2}}{T} \right) \\ &= O_p \left( \frac{\sqrt{m^3/l^2 + (ml)^3/T}}{T} \right) + O_p \left( \frac{\sqrt{l/T + 1/l^2}}{T} \right) + O_p \left( \frac{1}{Tm^{a-2}} \right). \end{aligned}$$

It then follows that  $\sup_{x \in [0, +\infty)} |P(F_T(\infty) \leq x) - P(F_T^*(\infty) \leq x)| \leq \sup_{x \in [0, +\infty)} |\psi_T(x) - \psi_T^*(x)| + o_p(1/T) = o_p(1/T)$ .  $\diamond$



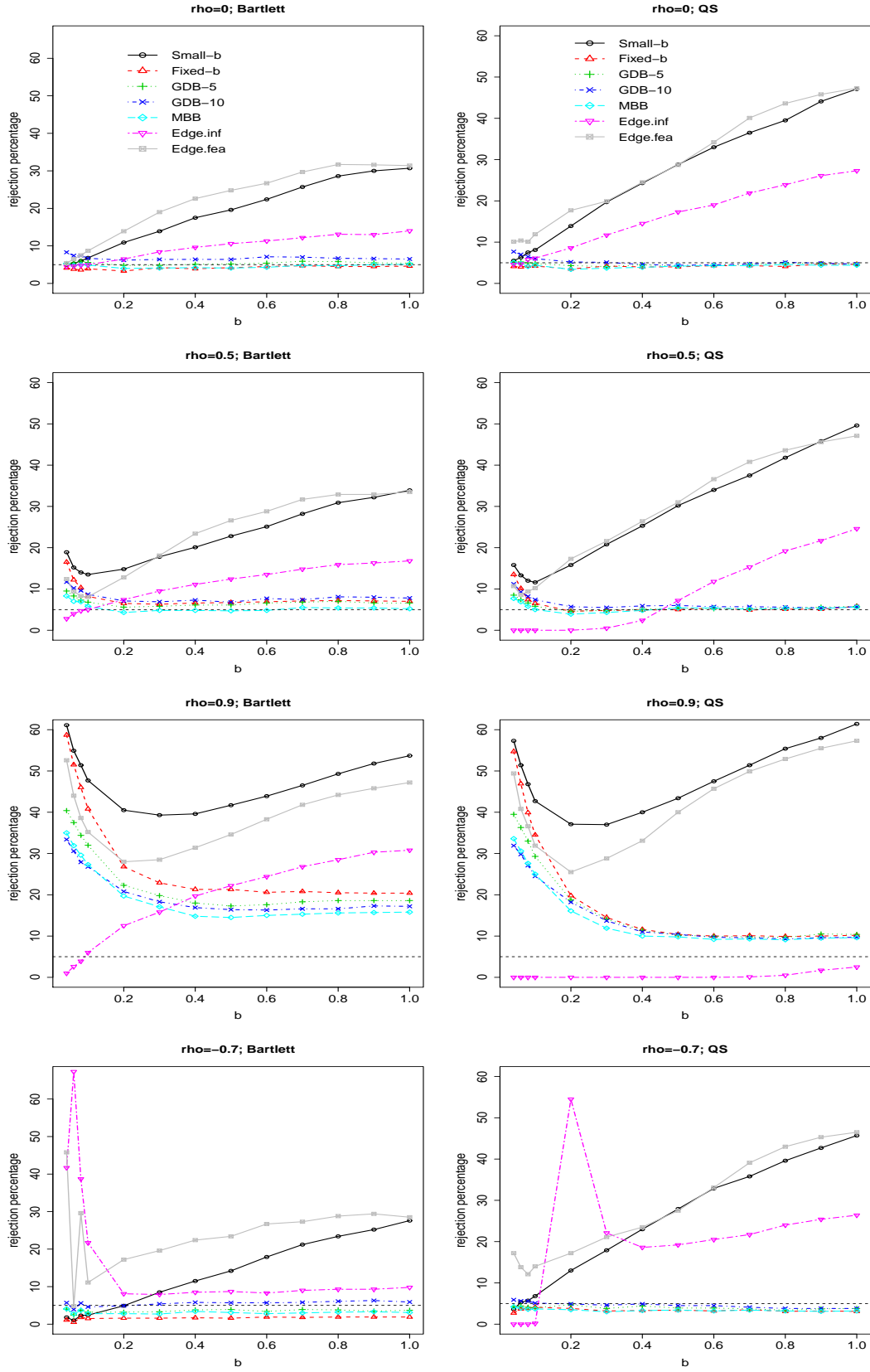


Figure 2: Empirical rejection probabilities for the Wald statistic with the Bartlett kernel (left panel) and QS kernel (right panel) and for the AR(1) model with  $N(0, 1)$  innovations

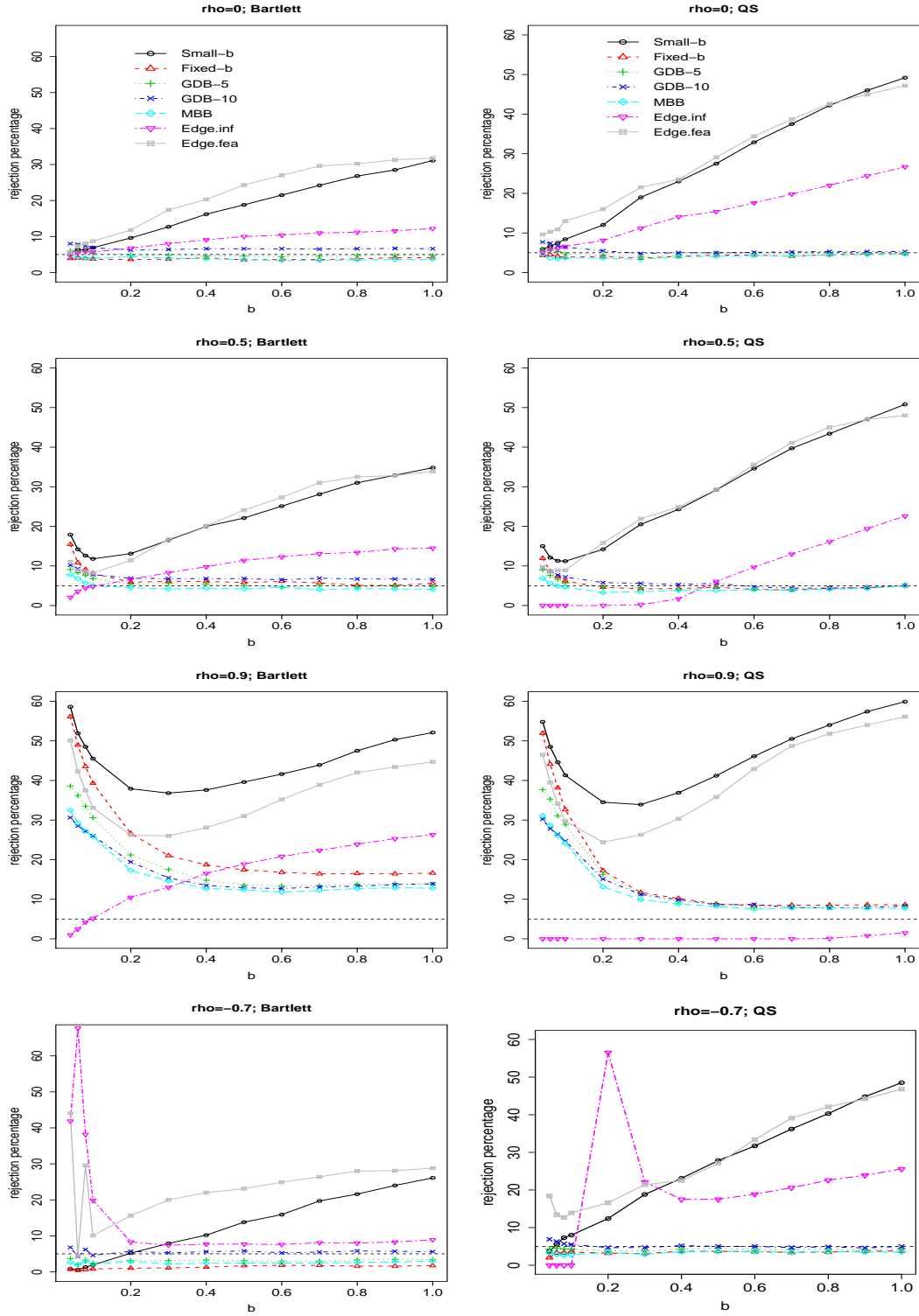


Figure 3: Empirical rejection probabilities for the Wald statistic with the Bartlett kernel (left panel) and QS kernel (right panel) and for the AR(1) model with  $t(3)$  innovations